kompendier
från

PEDAGOGISK-
PSYKOLOGISKA
INSTITUTIONEN

LÄRARHÖGSKOLAN
FACK, 200 45 MALMÖ 23

att strukturera olika datanivåer med hjälp av faktoranalystekniker

B. BIERSCHENK

Nr 12 - 1971

MEDDELANDE

FRÅN G F H

ATT STRUKTURERA OLIKA DATANIVÅER MED HJÄLP AV FAKTOR-ANALYSTEKNIKER

Bernhard Bierschenk

GRUPPEN FÖR FORSKNINGSMETODISK HANDLEDNING (GFH)

Nr 4, juni 1971

FÖRORD

Föreliggande redogörelse utgör ett försök att enkelt framställa ett svårt område. Hur detta har lyckats kan bara den enskilde läsaren avgöra. Att göra en s k elementär framställning av faktoranalys är i det närmaste en omöjlighet; det finns därför bara ett fåtal sådana försök. Man blir härvid tvingad att helt förbigå vissa delar inom faktoranalysen och dessutom ytligt behandla en del problem som man inte tycker sig kunna försumma. Författaren till detta meddelande från GFH har i stort sett begränsat sig till komponentanalys och en enkel (men vanlig) faktoranalysmodell. Även inom denna begränsning har en viss fokusering gjorts till nackdel för statistiskt hypotesprövande faktoranalys och olika rotationsmetoder. Förutom de tekniska svårigheterna har även tillgång till faktoranalytiska program vid Lunds Datacentral varit vägledande vid val av begränsningar. Ett tredje kriterium har antalet sidor utgjort, vilket haft till följd att utarbetade exempel på tillämpningar saknas. I gengäld har författaren kunnat ta med ett viktigt och sällan omtalat problem: strukturanalys av flerdimensionella datamatriser.

Bernt Larsson

INNEHÅLLSFÖRTECKNING

7104193

1. ANVÄNDNINGSMÖJLIGHETER FÖR FAKTORANALYSMETODER

Den påtagliga tillväxten och det tilltagande utnyttjandet av faktoranalytiska metoder är utan tvekan en framgång för denna analysteknik. Men samtidigt finns det tyvärr också många studier där faktoranalysen har använts på ett felaktigt sätt, kanske på grund av att för många faktor-analytiska läroböcker och artiklar har skrivits alltför tekniskt. Dessa framställningar är ofta förhållandevis svårlästa för den som vill bilda sig en uppfattning om olika analysteknikers egenskaper samt för- och nackdelar med hänsyn till speciella datauppsättningar.

I denna framställning ges därför en icke tekniskt hållen överblick över några problem som är förknippade dels med en faktoranalytisk bearbetning av olika datanivåer, dels med analysmodellens egenskaper. Syftet har ej varit att ge en mera systematisk och fullständig framställning. Ämnets komplexitet kräver för ett ingående studium en helt annan uppläggning. Den som vill fördjupa sitt vetande på det faktoranalytiska området hänvisas till läroböcker och artiklar om faktoranalysmetoder. Några dylika finns angivna i litteraturförteckningen.

Ett faktoranalytiskt problem börjar vanligen med förhoppningen eller övertygelsen att ett speciellt område inte alls är så kaotiskt som det vid första anblicken ser ut. Faktoranalytiska metoder utvecklades i syfte att i första hand identifiera huvuddimensioner eller kategorier med hänsyn till kognitiva funktioner. Olika metoder utvecklades för att studera individuella differenser bland människor. Dessa differenser borde dock betraktas som ett steg i riktning mot studiet av bakomliggande processer.

En faktoranalys kan endast utföras om de individer, som ingår i undersökningens population, skiljer sig från varandra.

I allmänhet urskiljes tre användningsformer: (1) att reducera data, (2) att söka efter en struktur och (3) att testa hypoteser.

Faktoranalysen är ej begränsad till antagandet att faktorerna är av fysiologisk eller sociologisk natur, av enkel eller komplex natur och inte heller att de är korrelerade eller okorrelerade.

Vissa faktorer kan nämligen t ex definieras som (1) endokrinologiska effekter, (2) biokemiska eller biofysiska parametrar, (3) neurofysiologiska eller vaskulära relationer på något anatomiskt område eller (4) dynamiken i det autonoma nervsystemet. De kan också definieras som (5) erfarenheter och skolning.

Faktoranalysen är särskilt användbar på de områden, där det ej finns redan etablerade begrepp, som skulle kunna tillåta utförandet av

experimentella prövningar. Faktorerna måste troligen ses som variabelgrupperingar, inom vilka vissa variabler kan vara mer eller mindre kraftigt relaterade till varandra. Dessa grupperingar utgör därför tämligen oberoende enheter.

Det skulle vara felaktigt att anta att faktorerna måste vara elementära. Betraktar man hela variabeluppsättningen inom en viss undersökning som representativ för ett visst område eller variabeldomän, som skall undersökas och är då syftet att undersöka om det finns en bakomliggande struktur eller ordning bland dessa variabler, så är detta ett faktoranalytiskt problem. Det faktoranalytiska problemet består då i att undersöka de observerade korrelationerna bland variablerna i syfte att upptäcka det minsta möjliga antalet faktorer med den minsta möjliga restvariansen (residualer).

Gränsen mellan olika användningsformer är flytande, men att skilja mellan dessa kan underlätta diskussionen. Hittills har faktoranalysen för det mesta använts för att upptäcka och säkra egenskapsstrukturer. Som regel kan sägas att de funna strukturerna är av tämligen komplex natur. Guilford antar t ex 120 faktorer som beskriver den intellektuella förmågan (se Child, 1970, s 73).

En allt vanligare användningsform som här skall behandlas något utförligare är: att reducera experimentella data till ett mindre, mera hanterligt antal variabler (faktorer). Denna användningsform av faktoranalystekniken förefaller kanske mindre dramatisk än det klassiska syftet att upptäcka allmänna karakteristika, men den är troligen icke mindre väsentlig. Har man t ex 50 eller 100 variabler (items) i ett frågeformulär eller skattningsschema vill man i en undersökning helst undvika att arbeta med alla dessa som beroende variabler. Det är i ett sådant fall önskvärt att gruppera variablerna på något sätt för att skapa möjlighet för en allmän överblick över undersökningens resultat. Grupperingen skulle man kunna åstadkomma med hjälp av faktoranalystekniken. Resultatet av en sådan bearbetning blir en sammansättning av ett antal items till en skala som motsvarar respektive faktor. Antalet skalor skulle t ex kunna bli 4 eller 5 och därmed mera lätthanterligt än 50 eller 100 variabler.

Ett mindre antal beroende variabler kan man också välja ut på annat sätt. Ett tillvägagångssätt skulle kunna vara att man i början väljer ut endast ett fåtal variabler. Problemet är då att man ej kan vara säker på att dessa variabler är representativa för variabeldomänen i fråga från vilken variablerna valts mer eller mindre godtyckligt. De för undersökningen utvalda variablerna kan t ex ha vissa specifika och oväntade egen-

skaper. Med en sådan teknik kan man även gå miste om en mätning av viktiga egenskaper som borde ha mätts.

Ett annat och bättre tillvägagångssätt skulle då ha varit att konstruera en a priori sammansatt skala där man utgår från logiska eller semantiska synpunkter. Men även i detta fall finns fortfarande risken att en sådan gruppering är missvisande. Mångdimensionalitet (tvetydighet) i vissa variabler skulle fortfarande kunna finnas kvar. Sådana tvetydigheter kan man upptäcka med faktoranalystekniken och sådana tvetydiga variabler kan man sedan utesluta från det fortsatta testkonstruktionsarbetet. Det är mycket svårt att göra samma sak i en a priori gruppering som har åstadkommits mer eller mindre av förnuftsskäl. En annan risk i en sådan gruppering är att man i olika sammanhang misslyckas med att konstruera två eller flera skalor som verkligen mäter samma allmänna faktor.

Att arbeta med enstaka variabler har också psykometriska nackdelar med hänsyn till den psykometriska felteorin. Vi är tvungna att gruppera ihop flera olika variabler för att åstadkomma stabila värden. I viss mån är detta mål naturligtvis uppnått genom a priori-grupperingsmetoden. Men faktoranalysen eller varje annan teknik som använder sig av den empiriska strukturen, är mindre beroende av en bra starthypotes med hänsyn till variablernas egenskaper.

Att använda faktoranalystekniken för att konstruera mätinstrument är i stort sett lika med datareduktionen.

Huvudskillnaden är att man har mera specificerade förväntningar på att vissa faktorer skall uppträda. En hel studie eller vissa delstudier kan vara utformade för att konstruera bra mätinstrument för mätning av vissa begrepp. Skillnaden mellan konstruktion av mätinstrument och upptäckande av vissa egenskaper är att man i det senare fallet har mindre klara föreställningar om vilka egenskaper som verkligen kan komma fram.

1.1 Faktorer som teoretiska konstruktioner

Syftet med faktoranalysen är att simultant identifiera flera olika obekanta. Ett grundläggande antagande i faktoranalysen är, att ett batteri av lineärt korrelerade variabler innehåller en allmän faktor, samt att individernas värden kan representeras mera ekonomiskt och överskådligt i form av referensfaktorer.

Men vad är egentligen en faktor? Faktorer har definierats som

(1) dimensioner, (2) determinanter, (3) funktionella enheter, (4) parametrar och (5) taxonomiska kategorier. Uttryckt i teoretiska termer har faktorerna betecknats som (1) klassificerande begreppsbildningar, (2) som verkliga ("real") och (3) som artificiella. Faktorerna har också beskrivits som (1) allmänna (generella) (2) grupperingar, (3) specifika och (4) som bipolara. Faktoranalysen har övervärderats och undervärderats. Man antog bl. a. att man med hjälp av faktoranalysen skulle kunna upptäcka invarianta faktorer, dvs faktorer som är demonstrerbara och upprepbara, om man bortser från vissa variationer i den ursprungliga (initiala) faktoriseringen av populationer, stickprov och mätinstrument.

Med faktor avses enligt J. R. Royce "en sann variabel, en process eller determinant som svarar för kovariationen i ett specificerat område, som de utförda observationerna avser".

Frågan i vilken mån däremot faktorerna kan anses vara operationellt definierade besvarar Royce så här: alla faktorer kan anses som operationellt definierade, eftersom faktorerna i varje fall uppträder induktivt som en funktion av flera olika specificerbara variabler. (Royce, 1967, s 323).

En faktoranalys börjar vanligen med en korrelationsmatris. Resultatet av en faktoranalys kan aldrig bli bättre än de data som analysen baseras på, dvs med vilka analysen påbörjas. De två steg som man då åtminstone först måste utföra är:

a. att undersöka frekvensfördelningen och
b. se till, att fördelningen ej blir påfallande snedfördelad, stympad eller multimodal.

Var och en av dessa avvikelser kan medföra risker för analysen. Om alla fördelningar är tämligen unimodala och symmetriska, är chansen stor att regressionen är lineär.

1. 2 Vad är en faktoranalys?

Faktoranalys är en statistisk teknik med vilken man simultant kan granska relationen mellan flera olika variabler eller variabeluppsättningar. Den relation som är av intresse för faktoranalysen är korrelationen, dvs man utnyttjar graden av överensstämmelse mellan olika variabler.

För det mesta har nämligen relationen mellan variabler som granskas simultant en systematisk struktur. Vanligen grupperar sig variabler i kluster med en hög överensstämmelse inom och låg överensstämmelse mellan klustren. Faktoranalystekniken bygger på en mycket specifik

matematisk modell för beskrivning av denna systematiska struktur. Med modellen specificeras på vilket sätt faktorerna bildar denna struktur. Strukturen användes sedan för att man skall kunna dra mera detaljerade slutsatser om faktorernas egenskaper och antal.

I allmänhet kan man med denna modell ej exakt återspegla de naturliga förhållandena, dvs modellen överensstämmer ej exakt med den empiriska datamängd som samlats in. Vi måste alltid vänta oss, att våra data medför restvarianser, troligen slumpmässiga variationer, som ej kan förklaras med modellen. Dessa variationer anses vara mätnings- eller samplingsfel.

Sådana fel medför att våra slutsatser från empiriska data till attribut (egenskaper, relationer) och antalet faktorer alltid är osäkra: Vi kan t ex aldrig få ett exakt svar på frågan hur många faktorer det finns i materialet och vilka egenskaper dessa faktorer har. Faktoranalysen skiljer sig alltså vid användning på en datamängd ej från andra matematisk-statistiska modeller. Det enda man kan hoppas på är att osäkerhetsgraden för inferenserna kan hållas så låg som möjligt.

Faktoranalysmodellen gör mycket starka antaganden om faktorernas natur. Om dessa antaganden på något sätt invalideras, kan resultatet av en faktoranalys ge en felaktig och vilseledande uppfattning om datamängdens struktur.

Ett syfte med faktoranalystekniken är att bestämma antalet faktorer. Ett annat syfte är att bestämma faktorladdningar och faktorvärden. Faktorladdningar avser variabelegenskaper. De anger hur väsentlig en speciell faktor är för de värden som observeras i en speciell variabel. Faktorladdningar kan också ses som korrelationer mellan faktorer och variabler. OBS: det gäller ej den vanliga korrelationsformen, eftersom en faktor är en konstruktion, dvs en latent variabel som, så som den är definierad, ej kan observeras på samma sätt som en manifest variabel. Faktorvärden beskriver individer och är liksom faktorladdningar en mätning av latenta variabler.

Faktorladdningar är variabelkarakteristika och faktorvärden individkarakteristika. Modellen antar att det inte finns någon interaktion mellan variabler och individer, en förutsättning som inte alltid kan anses vara uppfylld. Det antas att en viss variabel har samma uppsättning laddningar för alla individer och att en viss individ har samma uppsättning faktorvärden för alla variabler.

Faktoriseringen av en given korrelationsmatris med någon av de existerande faktoriseringsmetoderna medför att faktorlösningen inte är unik.

Axlarna roteras i syfte att förändra dem i förhållande till den första lösningen som bestämts genom extraheringsmetoden i riktning mot en position, som är användbar för en meningsfull tolkning av faktorerna samt för jämförelser med andra studier.

Huvudsyftet med en sådan rotation är att erhålla meningsfulla faktorer som dessutom med hänsyn till olika analyser är så konsistenta eller invarianta som möjligt.

Den mest lämpliga lokaliseringen eller placeringen av denna referensram är beroende av syfte och målsättning samt de teoretiska antaganden, som gäller för en bestämd undersökning.

Endast efter rotation med hänsyn till de olika frågeställningar eller hypoteser som gäller för ett visst område, som skall granskas, kan man vänta sig att få en viss tolkbarhet av faktorerna.

Medan en mekanisk rotation matematiskt är förhållandevis enkel, innebär bestämningen av ett kriterium för rotationen stora problem. Rotationen beror då i stort sett på hur man vill tolka sina data. Den är på så sätt baserad på den teoretiska utgångspunkten för en viss undersökning, resp på den enskilde forskarens betraktelsesätt.

1.3 Faktoranalysmodeller

Utvecklingen av den faktoranalytiska forskningen har lett till flera olika modeller. Faktoranalysen har utvecklats i samband med korrelationsberäkningar inom psykologin. Den äldsta modellen är Spearmans (1904) g-faktormodell. I denna modell spaltas varje variabel upp i två okorrelerade komponenter. Den ena komponenten motsvarar en för alla variabler gemensam grundfaktor kallad "allmän faktor" ("general factor") eller g-faktor. Den andra komponenten motsvarar en "specifik faktor" ("specific factor"), som endast kan hänföras till en enda bestämd variabel. Variablernas korrelationer måste man i denna modell kunna reproducera enbart ur g-faktorns laddningar, om de specifika faktorerna verkligen är okorrelerade med varandra.

Även om denna modell i dag kan betraktas som felaktig, har den dock varit grundläggande för den faktoranalytiska utvecklingen.

En orsak till att man frångick Spearmans modell var att allt flera korrelationsmatriser publicerades, som ej kunde förklaras genom en

enda gemensam faktor. Det var först sir Cyril Burt (1909) som av denna orsak hårt kritiserade modellen och samtidigt utvecklade en faktoranalysmodell, som blev känd under beteckningen "gruppfaktormodell". Utöver de specifika faktorerna antar denna modell flera allmänna faktorer i stället för en enda.

Av Louis L. Thurstone (1931) presenterades en faktoranalysmodell som möjliggjorde analysen av flera gemensamma faktorer ("multiple factor analysis"). Denna modell blev den mest använda i den psykologiska forskningen. Den är en vidareutveckling av Burts gruppfaktormodell.

Oberoende av varandra utvecklades ungefär samtidigt av Harold Hotelling (1933) och av Truman L. Kelley (1935) en faktoranalysmodell som kallas huvudkomponent- eller principalkomponent- ("principal component") modellen.

Denna modell, som i dag förknippas framför allt med Hotellings namn, gör ingen skillnad mellan flera gemensamma (allmänna) och attributspecifika faktorer. Trots att modellen har formella fördelar framför andra mera speciella faktormodeller, har den bara i ringa utsträckning kommit till användning i psykologin.

Däremot har den av Hotelling föreslagna huvudaxel- eller principalaxel-metoden blivit huvudmetoden för att beräkna faktorerna.

Medan under senare tid Guttman vidareutvecklade faktoranalysens strukturteoretiska användningsmöjligheter, kunde Tucker presentera en avsevärd utvidgning av faktoranalysmodellen för multikategoriala datauppsättningar.

I fråga om innehåll var Raymond B. Cattell den förste, som presenterade ett systematiskt program för att med faktoranalytiska metoder undersöka individuella skillnader inom personlighetsforskningen. Det nya i Cattells insats är hans försök att åstadkomma en representativ registrering av alla personlighets-, temperaments-, motivations- etc variabler genom en samtidig undersökning av individen med hjälp av (1) självskattningar, (2) skattningar genom andra och (3) objektiva personlighetstest.

Medan Cattells försök innebär åstadkommande av representativitet, försökte Hans J. Eysenck en faktoranalytisk bearbetning av kliniskdiagnostiskt relevanta personlighetsattribut.

Guilfords faktoranalytiska undersökningar under senare tid gäller framför allt intelligens- och begåvningsforskning.

Utan att närmare gå in på dessa olika faktoranalytiska metoders för- och nackdelar kan det sägas, att den kanske enklaste och mest eleganta

av alla existerande faktoranalytiska metoder för att få faktorladdningsmatriser är principalaxelmetoden. Denna metod har betecknats som "principal-axel-lösning", "egenvärde-egenvektor-metod", "principalkomponent metod", "den karakteristiska rot- och vektormetoden" samt "den latenta rot- och vektormetoden". Thurstones utveckling av centroidmetoden är en approximering till principalkomponentmetoden som utvecklades av Hotelling.

1.4 Faktoranalys som en lineär flerdimensionell modell

Begreppet "faktoranalys" avser en flerdimensionell modell, som tillåter en analys av kvantitativa attribut (egenskaper, relationer) i funktionellt enklare bas- eller grundvariabler eller "faktorer". Ett faktoranalytiskt förfaringssätt är bundet till vissa bestämda multivariata undersökningsplaner. En rad olika förfaringssätt för dataanalysen kan härledas ur faktoranalysen. I en mindre strikt användning av begreppet "faktoranalys" avses för det mesta enbart sådana dataanalysmetoder. En mera strikt tillämpning av begreppet avser dock modell, undersökningsplan och analysmetod.

Faktoranalys som metod skulle kunna beskrivas på följande sätt: Inom ramen av de förutsättningar, som införs med denna modell, tillåter denna analys att bestämma antal och beskaffenhet av grundfunktioner, grundvariabler, beteendedimensioner eller faktorer, som är en funktion av de beteendefunktioner, som blev föremål för analysen.

Inom en undersökningsplan, vars syfte är att studera basfunktioner, på vilka de observerade beteendena bygger, utgör faktoranalysen en central del. Men för den skull kan faktoranalysen dock under inga förhållanden betraktas som en universalmetodologi. Inom ramen av en sådan undersökningsplan utgör faktoranalysen nämligen enbart ett av flera olika steg, nämligen: en dimensionsanalys, avseende en viss bestämd datauppsättning. Faktoranalysen är alltså en strukturanalys. Som ett nödvändigt komplement tillkommer i många undersökningar variansanalytiska modeller för att kunna utföra nivåanalyser.

När man talar om faktoranalys måste man urskilja en "allmän faktormodell" och de "specifika faktormodeller" som har härletts ur den allmänna modellen. Den allmänna faktormodellen definierar med hjälp av några få grundekvationer modellen för uppspaltning av variansen av en kvantitativ egenskap i ej direkt observerbara grundvariabler.

För att säkerställa, att faktorlösningarna är entydiga, behövs det därutöver ett antal tilläggsantaganden. Beroende på valet av vissa bestämda tilläggsantaganden får man en av många speciella faktormodeller.

I en faktoranalys utgår man från antagandet att varje kvantitativt attribut (egenskap eller relation) x kan uppfattas som en funktion av flera faktorer $f_1, f_2, \ldots, f_r$. Detta kan uttryckas som

$$x = \varphi(f_1, f_2, \ldots, f_r). \qquad (1)$$

I denna ekvation är enbart x känd, okända är såväl antalet faktorer som faktorernas funktionstyp φ. Dessa två moment måste bestämmas, om man vill definiera de för olika attribut gemensamma faktorerna. I faktoranalysen använder man sig av den matematiskt enkla funktionen, att x skall vara en linjär funktion av f, dvs en vägd summa av faktorerna. Ekvationen (1) kan då mera exakt skrivas som

$$x = a_1 f_1 + a_2 f_2 + \ldots a_r f_r \qquad (2)$$

Denna ekvation är lineär med hänsyn till f, och x är summan av alla f, var och en multiplicerad med den tillhörande koefficienten (vikten) a.

Faktorerna kan betraktas som nya variabler. Ur varje observerat attribut kan för varje fp och för varje faktor beräknas ett mätvärde. På så sätt kan vi också för varje fp beräkna medelvärde och varians med avseende på en viss faktor. Dessutom kan det beräknas en korrelationskoefficient mellan två olika faktorer f_p och f_q. Detta leder till ett annat krav på den allmänna faktormodellen, nämligen: faktorerna i ekvationen (2) måste bestämmas på ett sådant sätt att faktorernas korrelationer genomgående är noll. Här skall ej beröras faktormodeller i vilka korrelationer mellan faktorerna är tillåtna. Dessa får man f ö alltid ur sådana faktorlösningar där faktorerna är okorrelerade (Pawlik, 1968, s 19). Uttryckt i en formel betyder detta

$$r_{f_p f_q} = 0; \; p \neq q = 1, 2, \ldots, r. \qquad (3)$$

Med detta krav är säkerställt att faktorerna - i motsats till de observerade egenskaperna - är ömsesidigt lineärt okorrelerade. Med dessa ovan givna fåtaliga antaganden är den allmänna faktormodellen beskriven i sina allmänna drag. (En mera precis bestämning samt en konkretisering av de mera speciella modellerna avses ej här.)

Ekvationen (2) betecknas i den faktoranalytiska litteraturen som "specification equation" för attribut x med avseende till r faktorer. Faktorernas koefficienter kallas "faktorladdningar" av attribut x som anger

eller bestämmer "strukturen", dvs uppbyggandet av x i faktorerna $f_1, \ldots, f_r$. Ju högre en koefficient a är, desto högre, säges det, är x "laddad i den motsvarande faktorn".

En faktor f_p betraktas som utan inverkan på prestationen av x när faktorladdningen a_p av x ligger omkring noll. Tvärtom har en faktor i stigande grad andel vid konstituerandet av en prestation x ju högre resp faktor laddar på x.

Faktorladdningar utgör egenskapsparametrar för faktormodellen. Sätter man i ekvationen (2) i stället för x in en annan egenskap så förändras i allmänhet koefficienterna a. Dessa koefficienter är alltså EGENSKAPSSPECIFIKA, men samtidigt icke personspecifika. För alla fpp i en undersökning förutsättes samma "bestämningsekvation" för egenskapen x. Individuella skillnader mellan personer kan man i en faktormodell ej avläsa i faktorladdningar. Dessa skillnader får däremot uttryck i de s k faktorvärdena (dvs mätvärden för personer i faktorerna $f_1, \ldots, f_r$). Faktorvärdena är alltså med andra ord icke egenskapsspecifika. De är däremot PERSONSPECIFIKA. Dessa värden karakteriserar varje enskild person, men ingår samtidigt oförändrade i bestämningsekvationen med hänsyn till samtliga variabler av en undersökning. Faktorvärdena kallas därför också faktormodellens personparametrar. I stora stickprov borde faktorladdningarna vara invarianta mot tillfälliga variationer i personstickprov. Faktorvärdena å andra sidan borde vara invarianta mot tillfälliga variationer i egenskapsstickprov.

Att hålla isär egenskaps- och personparameter inom faktoranalysmodellen är av avgörande betydelse. På så sätt blir det nämligen möjligt att med hjälp av faktorladdningar framställa en allmän (för samtliga fpp gällande) regelbundenhet i den psykologiska strukturen för ett visst attribut. Samtidigt ger faktorvärdena möjlighet att göra individuella skillnader i samma beteendekomponenter mätbara.

Den allmänna bestämningsekvationen (2) för faktoranalysmodellen har följande grundstruktur ("basic structure"), se f ö Horst (1963, ss 81-82). Med "basic" avses en matris vars rang = min (m, n).

$$
\underset{m \text{ variabler}}{\overset{n \text{ Personer}}{\begin{bmatrix} x_{11} \ldots x_{1j} \ldots x_{1n} \\ \ldots \\ x_{i1} \ldots x_{ij} \ldots x_{in} \\ \ldots \\ x_{m1} \ldots x_{mj} \ldots x_{mn} \end{bmatrix}}} = \underset{m \text{ variabler}}{\overset{r \text{ Faktorer}}{\begin{bmatrix} a_{11} \ldots a_{1p} \ldots a_{1r} \\ \ldots \\ a_{i1} \ldots a_{ip} \ldots a_{ir} \\ \ldots \\ a_{m1} \ldots a_{mp} \ldots a_{mr} \end{bmatrix}}} \times \underset{r \text{ faktorer}}{\overset{n \text{ Personer.}}{\begin{bmatrix} f_{11} \ldots f_{1j} \ldots f_{1n} \\ \vdots \\ f_{p1} \ldots f_{pj} \ldots f_{pn} \\ \vdots \\ f_{r1} \ldots f_{rj} \ldots f_{rn} \end{bmatrix}}}
$$

$$X = A \quad F$$

Figur 1. Faktoranalysmodellens allmänna bestämningsekvation i matrisform.

där definitionen för X (observationsvärdematris)

$$x_{ij} = a_{ip} f_{pj} \tag{4}$$

a_{ip}: faktorladdningen för variabel x_i i faktor f_p.

Den fullständiga laddningsvektorn är definierad som

$$a_i = (a_{i1}\ a_{i2} \ldots a_{ip} \ldots a_{ir}). \tag{5}$$

Definition för F (faktorvärdesmatris)

$$
F = \begin{bmatrix} f_1 \\ \\ f_2 \\ \ldots \\ f_p \\ \ldots \\ f_r \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} & \ldots & f_{1j} & \ldots & f_{1n} \\ \\ f_{21} & f_{22} & \ldots & f_{2j} & \ldots & f_{2n} \\ & & & \ldots & & \\ f_{p1} & f_{p2} & \ldots & f_{pj} & \ldots & f_{pn} \\ & & & \ldots & & \\ f_{r1} & f_{r2} & \ldots & f_{rj} & \ldots & f_{rn} \end{bmatrix}
$$

Definition för A (faktormatris).

$$
A = \begin{bmatrix} a_1 \\ a_2 \\ \ldots \\ \\ a_i \\ \ldots \\ \\ a_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1p} & a_{1r} \\ a_{21} & a_{22} & \ldots & a_{2p} & a_{2r} \\ & & \ldots & & \\ \\ a_{i1} & a_{i2} & \ldots & a_{ip} & a_{ir} \\ & & \ldots & & \\ \\ a_{m1} & a_{m2} & \ldots & a_{mp} \ldots & a_{mr} \end{bmatrix}
$$

Symbolfunktioner

För att hålla symbolerna överskådliga använder man sig ofta av dubbla indices. Man kan utgå från det allmänna fallet, att n fpp i m < n kvantitativa variabler $x_1, x_2, \ldots, x_m$ undersöks, vars fullständiga faktorlösning medför $r \geqq m$ faktorer $f_1, f_2, \ldots, f_r$ (där alltid $r < n$). Som index användes för

variabler: $g \neq i = 1, 2, \ldots, m$,

personer: $h \neq j = 1, 2, \ldots, n$,

faktorer: $p \neq q = 1, 2, \ldots, r$,

där

x_{ij} = mätvärde för j:te fp, i:te attribut

f_{pj} = mätvärde för j:te fp, p:te faktor

Formler kan ofta skrivas enklare i matris- och vektorform. Därvid betecknar x_g resp x_i den n-dimensionerade radvektorn för mätvärden för n fpp i attribut x_g resp x_i. Det summerade antalet m variabelvektorer $x_1, x_2, \ldots, x_m$, skrivna under varandra, resulterar i datamatrisen X av typen (m, n), där m n experimentella observationer ligger till grund för faktoranalysen. När man använder sig av vektorbegrepp för att uttrycka faktorer, skriver man f_p resp f_q för den n-dimensionerade radvektorn för n fpp:s värden i faktorn f_p resp f_q.

Variabelvektorn x_i är en lineärkombination av de r lineära oberoende vektorerna $f_1, \ldots, f_r$; Koefficienterna $a_{i1}, \ldots, a_{ir}$ av denna lineärkombination kallas faktorladdningar till attribut i för de r faktorerna.

1. 5 Att skilja mellan faktormodell, faktorextraheringsmetoder och faktorrotationsmetoder

Grunddefinitionen av faktormodellen görs alltså med en lineär regression av varje manifest variabel på de latenta variablerna eller faktorerna. En faktormatris uttrycker dessa lineära kombinationer. Elementen i denna matris är lineära regressionsvikter för de manifesta variablernas regression på de latenta variablerna. I faktormodellen förklarar de k gemensamma faktorerna den gemensamma variansen i datamatrisen. Specifik varians och felvarians betraktas för det mesta som en residualterm, dvs som en felkomponent.

De metoder, som medför att man inom ramen av de begränsningar som gäller för en bestämd faktormodell kan extrahera faktorer, kallas faktorextraheringsmetoder. Att extrahera faktorer avser en teknik som

tillåter att man ur de manifesta variablernas korrelationsmatris kan beräkna en faktormatris. Spearmans g-faktormodell och Lawleys maximum-likelihood-metod föreskriver modellspecifika extraheringstekniker. Men principalkomponentmetoden och modellen för flera gemensamma faktorer föreskriver inga begränsningar.

Mellan de två sistnämnda modellerna är extraheringsmetoderna utbytbara. Bortser man från Harmans diagonalmetod och gruppfaktoranalysen finns det två modelloberoende faktorextraheringsmetoder som man kan välja mellan: (1) principalaxelmetoden och (2) den något grövre och ur matematiskt-tekniska synpunkter enklare centroidmetoden. Principalaxelmetoden är av intresse ur följande synpunkter:

1. Varje faktor extraherar maximal varians (t ex kvadratsumman för faktorladdningarna är maximerade för varje faktor) och ger på så sätt de minsta möjliga residual-matriserna.
2. Korrelationsmatrisen krymps genom denna metod till det minsta möjliga antalet ortogonala faktorer.
3. Metoden bygger på minsta kvadratmetoden för en given korrelationstabell.
4. Hotelling (1936) utvecklade ett iterativt förfaringssätt för att erhålla faktorladdningarna för principalaxlarna, vilket uppnår en tillfredsställande exakthet för en viss önskad nivå.

Utvecklingsmässigt hör dock principalaxelmetoden till principalkomponentmodellen. Centroidmetoden utvecklades i samband med modellen för flera gemensamma faktorer. Faktorerna kan man naturligtvis åstadkomma genom att denna bas roteras med vilken ortogonal rotation som helst. Först genom tilläggskravet, som kommer till uttryck i faktorextraheringsmetoden och som ställs på faktorerna, fastställer man faktorernas position. Men denna bestämning är villkorlig med hänsyn till modellen och vad som är mera väsentligt, ej nödvändigtvis psykologiskt tillfredsställande. Vore det enda syftet att representera den funktionella relationen mellan variablerna (strukturmodell) resp en reduktion av utgångsdata, dvs k $<$m ursprungsvariabler (datareduktionsmodell), spelar frågan, på vilket sätt man placerar koordinataxlarna i faktorrymden, ingen roll.

Situationen blir dock helt annorlunda, när faktorerna skall tolkas innehållsmässigt, eftersom varje rotation medför olika tolkningsmöjligheter. Matematiskt existerar det ett oändligt antal faktormatriser som reproducerar gemensam varians i materialet. Genom en rotering av faktorerna runt origo kan man i princip få ett oändligt antal nya

system av dimensioner. Detta kan ske med eller utan bibehållande av ortogonaliteten (rätvinkligheten).

Vid användning av faktoranalysmodeller söker man efter en psykologiskt meningsfull struktur, dvs efter tolkbara faktorer. Under den senaste tiden har man försökt utveckla matematiska kriterier för att komma fram till meningsfulla dimensionsuppsättningar. Detta har lett till en mängd olika s k rotationsförfaranden.

Redan i sin andra uppsats till teorin för flera gemensamma faktorer hänvisade Thurstone till problemet att den direkta faktorlösningen enbart ger ett första relationssystem vars koordinataxlar (dvs faktorer) först måste roteras efter vissa kriterier, innan faktorlösningen blir psykologiskt meningsfull, tolkbar och generaliserbar.

För rotationstekniken utgör faktormatrisen basen. En faktors markeringsvariabler, dvs de variabler som har höga laddningar i denna faktor, ger en anvisning om vilka beteendeegenskaper som i första hand definieras av denna faktor. På grundval av detta kan man sedan härleda hypoteser om faktorns innehåll, dvs den psykologiska betydelsen (för olika rotationstekniker se kap. 9).

Sammanfattningsvis kan sägas: Faktoranalysen ger basen för faktorrymden, dvs fullständigt antal, r, ömsesidigt ortogonala vektorer som representerar denna rymd och inom vilken variablerna kan presenteras. I en ekvivalent representation av faktorladdningar som på grund av detta resulterar i en annan faktortolkning, blir rotationsfrågan av central betydelse.

Inom psykologin används i de flesta faktoranalytiska undersökningarna Thurstones enkel-struktur-rotationskriterium. Syftet är här att åstadkomma någon form av transformering av den oroterade matrisen på ett sådant sätt, att den nya faktorladdningsmatrisen får följande karakteristika:

1. För varje faktor skall det enbart finnas ett relativt litet antal variabler med höga laddningar. De resterande variablerna skall ha små laddningar.
2. Varje variabel får enbart ha höga laddningar i ett fåtal faktorer.
3. För vilket faktorpar som helst skall ett stort antal variabler ha endast små laddningar med hänsyn till båda faktorerna.
4. För ett givet faktorpar skall några variabler ha höga laddningar i en av faktorerna och låga laddningar i den andra faktorn, medan andra variabler har höga laddningar i den andra faktorn, men låga laddningar i den första.

5. För ett givet faktorpar får det endast finnas ett mycket litet antal variabler, som skulle kunna ha höga laddningar i båda faktorerna.

Dessa fem betingelser formulerades av Thurstone (1947) som enkelstruktur-kriteriet. Faktorrotationer, som har till syfte att uppnå dessa krav, kallas enkelstruktur-transformationer och den transformerade faktorladdningsmatrisen kallas en enkelstruktur-faktorladdningsmatris (Horst, 1965, s 385).

Men frågan om enkel-struktur-rotationen också medför en entydig enkel-struktur kan ej besvaras. Hittills har man ej kunnat visa att det till en given faktormatris endast finns en enda roterad matris som bäst motsvarar Thurstones enkel-struktur-kriterium. Thurstone menar därför att först en faktors replicerbarhet, dels med hänsyn till ett nytt personstickprov, dels med hänsyn till ett nytt attributstickprov, kan anses som beviskriterium (Pawlik, 1968, ss 179-180).

2. ATT FAKTORANALYSERA OLIKA DATANIVÅER

Den mest använda faktoranalystekniken är den s k R-tekniken, dvs man faktoriserar variabler med denna teknik. Byter man ut personer mot variabler och variablerna mot personer, får man en faktoranalys, som bygger på den s k Q-tekniken. Denna teknik betecknas också som den till R-tekniken hörande inversa faktoranalysen.

Med hjälp av Q-tekniken undersöker man n personer i vissa egenskaper och faktoranalyserar personer i stället för variabler. Medan R-tekniken utgår från en "egenskaps"-korrelationsmatris, bygger Q-tekniken på en korrelationsmatris, som är resultatet av korrelationen mellan personer. (I regel standardiseras för detta syfte varje variabel över personer).

Med hjälp av R-tekniken får man egenskapsfaktorer. Q-tekniken resulterar däremot i personfaktorer eller "typer". Faktormatrisen ur en Q-analys innehåller faktorladdningarna för personerna i de enskilda typfaktorerna. Dessa typfaktorer ger upplysning om i vilken utsträckning en viss fp tillhör en bestämd typ. Omvänt gäller, att faktorvärdematrisen ur en Q-analys innehåller standardvärdena för variablerna i de olika typfaktorerna. Ur dessa värden kan man avläsa i vilken riktning en egenskap är relevant för en bestämd typ. I Q-analyser undersöker man ett stort antal variabler hos ett relativt litet antal personer. Användningsområdet för denna analysform ligger i huvudsak inom den kliniska forskningen.

Att R- och Q-tekniken resulterar i samma gemensamma faktorer

är enbart påvisbart för den vanliga faktoranalysen, när datamatrisen såväl radvis (dvs över fpp) som kolumnvis (över variablerna inom varje fp) standardiseras, dvs när man genomför en dubbelstandardisering av datamatrisen (Pawlik, 1968, s 280).

Tillsammans beskriver R- och Q-tekniken dock främst en av flera möjliga "psykologiska datanivåer".

Med en psykologisk datanivå avses en matris av observationsresultat, vars rader och kolumner består av två olika kategorier. I en R- och Q-teknik är observationskategorierna försökspersoner och observerade egenskaper.

Därutöver finns dock ytterligare en rad andra psykologiska iakttagelser, som i sin tur medför nya datanivåer. Exempel på sådana är tidpunkt för mätningar, situationer, i vilka mätningen utförs, bedömarens personkarakteristika och olika undersökningsmetoder (självvärdering, värdering genom andra, objektiva test).

Trots att man kan klassificera olika beteendeobservationer i flera olika kategorier, har man i den klassiska faktoranalysen vid varje tillfälle begränsat sig till två kategorier. Alla andra har man hållit konstanta.

Cattell (1946) hänvisade till möjligheten att utvidga faktoranalysen på andra än person-variabel-datanivå, R-Q-tekniken, samt beskriver sådana analystekniker. Senare har Coan (1961) byggt ut Cattells schema.

Cattell tog hänsyn till tre kategorier, avseende psykologiska data: (1) observerad egenskap, (2) försöksperson och (3) observationssituation. Dessa fick beteckningarna V (variabel), P (person) och S (situation). Varierar man alla tre samtidigt, dvs undersöker olika personer under olika betingelser med samma ordningsföljd av egenskaperna, formas av dessa data ett tredimensionellt koordinatsystem med axlarna V, P och S. Dessa tre kategorier medför tre olika datanivåer av typen V-P, V-S och P-S. Dessa typer skiljer sig från varandra, beroende på de kategorier, i vilka man ej varierar data. På nivån V-P bortser man från dimensionen S, dvs alla observationer utförs under samma betingelser (i samma standardiserade försökssituation). På nivån V-S bortser man från skillnaderna mellan personer (analysen begränsas till en enda person) och på datanivå P-S håller man kategorien V konstant (analysen begränsas till en enda variabel).

På varje nivå kan faktoranalysen genomföras med hjälp av två alternativa tekniker. På nivån V-P t ex kan man genomföra antingen en R- eller Q-analys, beroende på vilket sätt man beräknar korrelationerna

samt på vilket sätt man faktoranalyserar dessa.

Vid användningen av P-tekniken testar man en enda försöksperson upprepade gånger med samma variabeluppsättning. Dessa variabler korreleras och faktoranalyseras över situationer. Resultatet är s k tillståndsfaktorer ("state factors"), dvs faktorer som resulterar ur kovariationen över tid. Variabler, som är laddade i samma tillståndsfaktor, visar samma utveckling med avseende på de undersökta situationerna.

O-tekniker är den inversa analystekniken till P-analysen. I O-tekniken korreleras och faktoranalyseras situationerna över variablerna.

I undersökningar, i vilka man använder sig av T-tekniken, testas ett stickprov personer vid olika tillfällen i en och samma variabel. Man korrelerar situationerna över personer. Resultatet blir då s k "situationsfaktorer" för denna variabel. T-tekniken är en lämplig teknik t ex för att faktoranalytiskt undersöka utvecklingspsykologiska frågor samt för att undersöka inlärningsförlopp. Statistiskt innebär denna analysform en flerdimensionell analys av tidsserier.

S-tekniken är den till denna analys hörande inversa analysmetoden. Vid användning av S-tekniken korreleras och faktoranalyseras personer med avseende på en viss bestämd variabel över situationer.

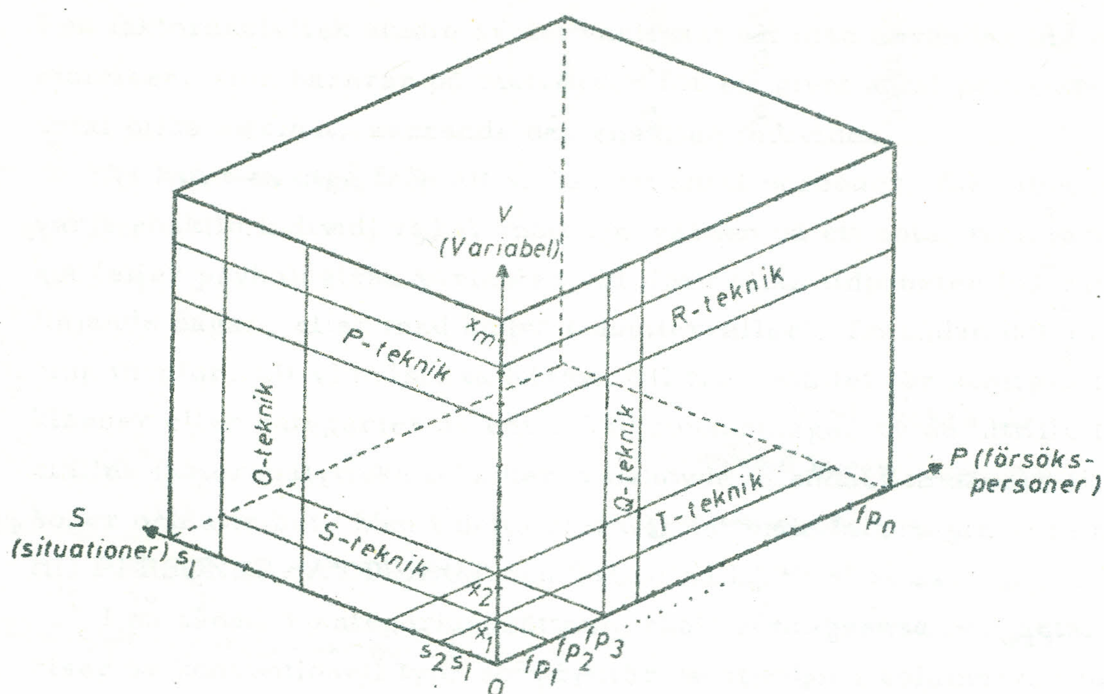
På samma sätt som R- och Q-tekniken är också P- och O-tekniken samt T- och S-tekniken reciproka med avseende till sina resultat.

Cattells kovariationsschema skulle redan genom användningen av R, P och T-tekniken kunna anses vara fullständigt utnyttjat. Men en sådan fraktionerad faktoranalys av en rätvinklig dataparallellipiped skulle medföra en tungt vägande nackdel: Eftersom vid varje fall en av kategorierna elimineras, kan man ej bestämma på vilket sätt faktorerna av alla de tre kategorierna hänger ihop. Horst (1963, 1965) utvecklade en metodik för att kunna simultant analysera tre- och flerdimensionella datauppsättningar. I avsnitt 3 redovisas Horsts metodik för ett tredimensionellt fall. Tucker (1963, 1967) lyckades vidareutveckla faktoranalysmetoden även på denna punkt genom att skapa en mera allmän faktoranalysmodell, som tar hand om flera än två kategorier.

Tuckers metod tar simultant hänsyn till egenskapsfaktor, personfaktorer och situationsfaktor. Dessutom ger denna analys upplysning om på vilket sätt faktorerna i dessa tre kategorier hänger ihop.

Metodiskt är denna faktoranalysteknik ännu ej tillräckligt utvecklad. Kommunalitetsskattningen och faktorrotationen med hänsyn till flera kategorier har hittills inte lösts tillfredsställande.

Figur 1 och tabell 1 sammanfattar vad som har anförts ovan.



Figur 2. Det något modifierade kovariationsschemat enligt Cattell 1946 (ur Pawlik, 1968, s 282).

Tabell 1. Grundläggande faktoranalytiska metoder enligt Cattell (ur Pawlik, 1968, s 289).

Data-nivå	Teknik	Kategorier som hålles konstant	Kategorier som interkorreleras och faktoranalyseras	Resulterande faktorer
V-P	R	Situation	Variabler	Egenskaps-faktorer ("trait factors")
	Q		Personer	Typfaktorer
V-S	P	Försökspers	Variabler	Tillstånds-faktorer ("state factors")
	O		Situationer	Situations-faktorer
P-S	T	Variabel	Situationer	Situations-faktorer
	S		Personer	Typfaktorer

3. ATT FAKTORANALYSERA MULTIPLA KATEGORIUPPSÄTTNINGAR

I en faktoranalytisk studie är det vanligast att man använder sig av datamatriser, som baseras på mätvärden för ett givet antal personer på ett antal olika attribut, avseende den enskilde individen.

Vi kan t ex utgå från att vi har ett antal personer, för vilka (för varje enskild individ) vi har mätt upp värden på ett antal fysiologiska och/eller psykologiska variabler vid flera olika tidpunkter (på varandra följande dagar, eller med andra tidsintervaller). En sådan datauppsättning medför, att vi måste ta hänsyn till tre i stället för vanligen två klasser eller kategorier av data. Vid användningen av de hittills framställda faktoranalytiska teknikerna behövde vi endast ta hänsyn till personer och attribut. Men i detta mera omfattande fall måste vi ta hänsyn till PERSONER, ATTRIBUT och TILLFÄLLEN.

I en sådan 3-kategoriuppsättning skall vi nu granska ett antal matriser av konventionell typ, som består av attribut i kolumnerna och personer i raderna. Varje enskild matris bör då bestå av en identisk attribut- och personuppsättning för ett antal olika tillfällen. Horst (1965, s 315) kallar dessa matriser skivor ("slabs") i fråga om 3-kategoridatamatrisen.

Horst (1963, ss 104-121) har diskuterat en rad allmänna problem kring en sådan uppsättning av data och föreslagit flera olika sätt att analysera en uppsättning av denna typ.

Även Tucker (1963, ss 122-137) diskuterade allmänna problem kring person-, attribut- och tillfälle-kategorier, som han betecknar som "the three-mode factor analysis model". Han presenterade enligt Horst (1965, s 316) ett skarpsinnigt förfaringssätt för att granska denna typ av datamatris.

Metoden förutsätter en lägre rang för varje enskild kategori, jämfört med de ursprungliga datamatriserna. Problemet består därvid i att av 3-kategorimatrisen med denna lägre rang, som bas, skapa reproduceringsmöjlighet för den observerade 3-kategori-datamatrisen (dvs ursprungsmatrisen).

Horst (1965, s 316) anser, att det också finns ett enklare förfaringssätt för att behandla data av denna typ, nämligen att man reducerar denna flerdimensionella datauppsättning till en 2-kategori-variabeluppsättning. Denna enklare väg skulle då innebära, att man analyserar data av denna typ med de konventionella faktoriseringsteknikerna.

Ett första sätt att behandla dessa data skulle kunna vara följande: Varje tillfälle betraktas för varje attribut som ett "distinkt attribut".

Detta synsätt skulle innebära att vi behandlar ett attribut, mätt vid första testtillfället, på ett stickprov personer som skilt från samma attribut, mätt vid nästa testtillfälle på samma stickprov personer.

För t ex fyra testtillfällen och 10 attribut läggs data ut så att en datamatris av typen personer gånger attribut resulterar i 40 olika attribut (fall 1).

Analysen utförs därefter med t ex principalaxelmetoden, där man får faktorladdningar för varje variabel (attribut) vid varje testtillfälle.

Ett annat sätt (fall 2) att behandla samma datauppsättning skulle kunna vara att betrakta varje testtillfälle som en från de andra skild uppsättning (stickprov) personer.

Om problemställningen behandlas på detta sätt, skulle datamatrisen kunna läggas ut, baserad på person-attribut-skivor för de olika testtillfällena, på följande sätt (vi antar 100 personer och fyra testtillfällen) 100 personer gånger fyra testtillfällen. Omstruktureringen resulterar därvid i 400 olika personer.

Fall 2 medför alltså att man baserar analysen på 400 personer och 10 attribut. Denna datamatris tillåter att bearbetningen av data sker med hjälp av en vanlig faktoranalysmetod. Denna resulterar i en uppsättning av faktorvärden för varje enskild person och för varje enskilt testtillfälle. Denna analysteknik skulle därvid resultera i faktorladdningar för de 10 variablerna.

Fall 1 och fall 2 medför principiellt olika resultat. I första fallet får vi en faktorvärdesmatris för 100 försökspersoner och en faktorladdningsmatris för de 10 attributen med hänsyn till varje enskilt testtillfälle.

Fall 2 medför en faktorvärdesmatris för försökspersonerna med hänsyn till varje enskilt testtillfälle och en faktorladdningsmatris, avseende de 10 attributen.

Sammanfattning:

Fall 1. Vi behandlar de olika testtillfällena som från varandra skilda variabler.

Fall 2. Vi behandlar de olika testtillfällena som olika personer.

Hittills har vi diskuterat upplösningen av en dataparallellipiped i testtillfälleskivor. Varje skiva fick personer i datamatrisens rader och attribut i dess kolumner. I fortsättningen behandlas nu en annan upplösningsteknik för en sådan 3-dimensionell datamatris, som medför att varje skiva representerar en person.

Varje personmatris skulle kunna betraktas som en matris, uppbyggd

med testtillfällen i raderna och attribut i kolumnerna. Om vi tillämpar denna matris på det ovan behandlade exemplet skulle vi få 100 matriser av ordningen 4 x 10.

Det finns även i detta fall 2 olika sätt att lösa problemet. Fall 1 innebär: att vi skulle kunna lägga ut skivor, så att varje person betraktas som ett från andra skilt attribut. Datamatrisens utseende skulle därvid vara följande: 4 testtillfällen och (10 x 100), dvs 1 000 attribut. Fall 2 innebär: att vi lägger ut matrisen i den andra riktningen och varje person betraktas som ett från andra skilt testtillfälle. Det skulle resultera i en matris med 400 rader och 10 kolumner.

Detta sätt att lägga ut data är identiskt med fall 2 ovan, dock med ett undantag: ett ömsesidigt utbyte av raderna har utförts. I båda fallen består kolumnerna av de 10 attributen, men individerna är för varje testtillfälle samlade, dvs grupperade i raderna.

I fall 2 har vi alltså samma 10 attribut, men testtillfälle-raderna har hopgrupperats med hänsyn till en speciell person.

Det har alltså hittills beskrivits tre olika möjligheter att lägga ut data i form av en tvådimensionell datamatris, som kan faktoranalyseras med tillgängliga metoder. Vad händer nu om vi använder det återstående tredje sättet att lösa upp data-person-testtillfälle-uppsättningen?

I den första typen av datamatris behandlades skivor på ett sådant sätt att varje skiva representerade ett av olika från varandra skilda testtillfällen.

I den andra typen av datamatris behandlades skivor så att varje skiva representerade en annan person.

I den tredje typen av datamatris behandlades skivor så att varje skiva representerar ett från andra skilt attribut.

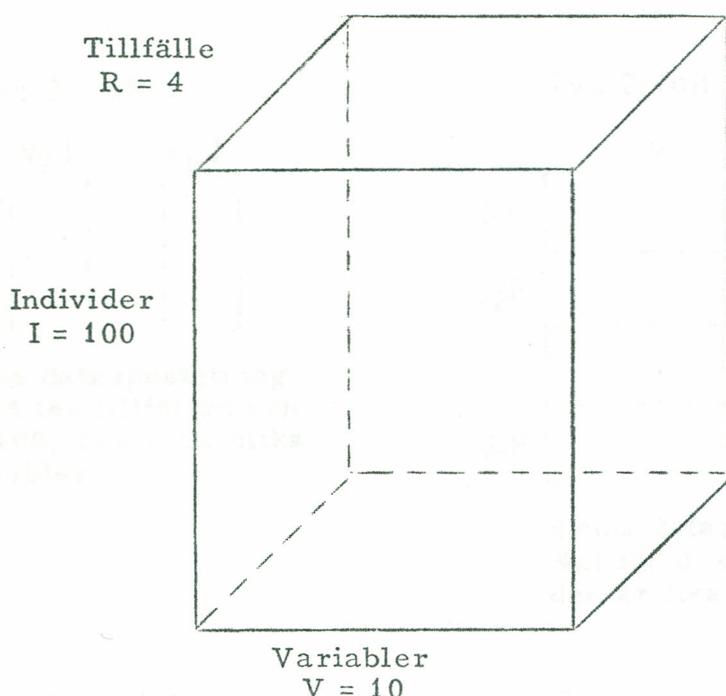
Nu ses varje skiva som en matris, som har personer i raderna och testtillfällen i kolumnerna. Om man nu lägger ut dessa skivor i form av en tvådimensionell datamatris, så att de 100 försökspersonerna konstituerar raderna och de olika uppsättningarna av testtillfällen bildar kolumnerna för varje av dessa successiva attribut - vad får man då?

Denna datauppsättning överensstämmer med den första datauppsättningen, som diskuterades ovan under attribut-personer-skivor, med undantaget att de olika testtillfällena nu är grupperade via attributen. I första fallet hade däremot attributen grupperats via de successiva testtillfällena.

Om man nu lägger ut data (person-testtillfälle-skivor) i en vertikal form, får man en datamatris, där kolumnerna består av testtillfällen och raderna innehåller personerna grupperade genom de successiva attributen.

Ett sådant arrangemang medför samma resultat som fall 1 vid attribut-testtillfälle-uppsättningen med ett undantag: Submatriserna ger en gruppering av personer i förhållande till de successiva attributen i stället för en gruppering av attributen i förhållande till successiva personer.

Genom dessa två omstruktureringar av data i person-testtillfälle-skivor har alltså egentligen ej tillkommit något nytt. Dessa olika förfaringssätt kan grafiskt åskådliggöras på följande sätt: Det tredimensionella koordinatsystem med 10 variabler på x-axeln och 100 individer på y-axeln samt 4 testtillfällen på z-axeln (figur 3) omstruktureras till 6 tvådimensionella datamatriser enligt figur 4, sid 27.

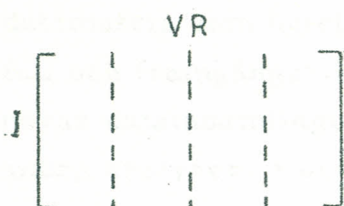


Figur 3. Parallellipiped (rätvinklig)

Samma resonemang skulle kunna tillämpas på försöksplan med fler än 3 kategorier av variabler. Men denna applicering skall ej behandlas här. (För diskussion av en sådan tillämpning, se Horst (1965, ss 318-324).

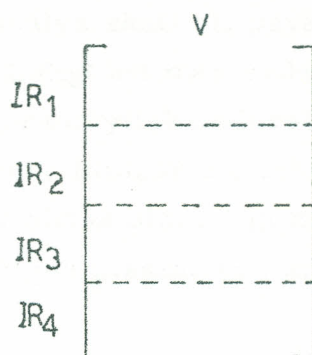
Denna mera utförliga framställning av behandlingen av multipla kategoriuppsättningar (se figur 4, sid 27) har gjorts, därför att en sådan analys måste vara av särskilt intresse för den psykologisk-pedagogiska forskningen. Problem, som är förknippade med denna analysteknik och utvecklingen av effektiva faktoranalysmetoder, som tar hand om multipla kategoriuppsättningar, har nämligen den senaste tiden alltmera kommit i uppmärksamhetens centrum. Syftet är att kunna reproducera datamatriser som innehåller multipla kategoriuppsättningar (3 och flera kategorier) med ett mindre antal parametrar. Detta innebär en generalisering av lägre rangapproximering till datamatriser med två kategorier.

Typ 1: fall 1



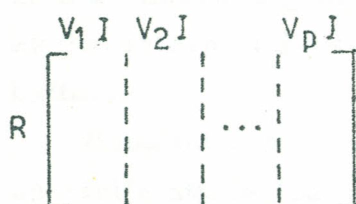
denna datauppsättning ger 4x10 dvs 40 olika variabler

Typ 1: fall 2



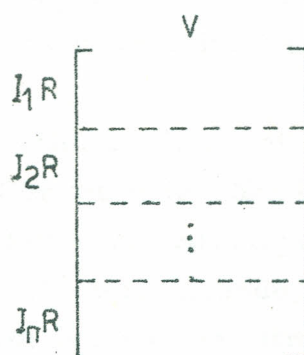
denna datauppsättning ger 100x4, dvs 400 olika personer

Typ 2: fall 1



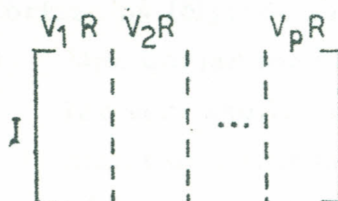
denna datauppsättning ger 4 testtillfällen och 10x100, dvs 1000 olika variabler

Typ 2: fall 2



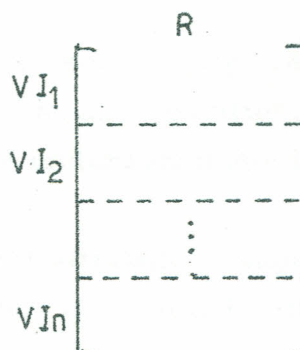
denna datauppsättning ger 4x100, dvs 400 olika personer, den är lika med Typ 1: fall 2

Typ 3: fall 1



denna datauppsättning ger 10x4, dvs 40 olika variabler, den är lika med Typ 1: fall 1

Typ 3: fall 2



denna datauppsättning ger 100x10, dvs 1000 olika personer, den är lika med Typ 2: fall 1

Figur 4. Sex tvådimensionella datamatriser

7104217

4. ATT PLANERA FAKTORANALYTISKA UNDERSÖKNINGAR

Rent matematiskt kan de faktoranalytiska metoderna användas på vilken datamatris som helst. För att en analys skall bli psykologiskt meningsfull och framgångsrik, är det nödvändigt att man redan från början planerar datainsamlingen som en faktoranalytisk undersökning. Liksom andra analyser, t ex variansanalyser, lämpar sig faktoranalyser dåligt som post mortem-analyser. Det är alltså olämpligt att använda denna analysteknik på data som ursprungligen insamlats i annat sammanhang och för andra syften.

4.1 Attributstickprov

Man skiljer mellan attributfixerade faktoranalyser och faktoranalyser av attributområden.

När det gäller attributfixerade faktoranalyser, avser man variabler, som är fixerade genom en speciell frågeställning, t ex vid undersökning av den interna faktorstrukturen i ett av flera deltest sammansatt testbatteri.

Resultaten av sådana analyser gäller i allmänhet enbart för denna speciella attributsammansättning. För att undersöka i vilken utsträckning sådana resultat kan generaliseras, krävs det ytterligare analyser.

Teoretiskt av större betydelse och mera givande, men också mera komplicerade, är faktoranalyser som avser olika attributområden. I sådana fall är i början enbart en mer eller mindre väl definierad attributpopulation ("variable domain") fastställd, t ex minnesprestation, socialt beteende eller skillnader mellan exempelvis normala och neurotiker. Ur en sådan population väljer man ut ett lämpligt stickprov. Har man att göra med ett nytt, hittills ej undersökt, attributområde, bör man förfara på följande sätt:

1. Man börjar med att så fullständigt som möjligt räkna upp alla variabler som skulle kunna räknas till denna population. Som källor kan man t ex använda sig av tidigare undersökningar inom besläktade områden, expertbedömningar etc.
2. När man på så sätt har definierat attributområden, väljer man ut ett lämpligt stickprov. Därvid bör man se till, att alla delområden blir representerade.
3. Sedan måste man utveckla lämpliga observations- och mätinstrument. Ibland kan man använda sig av en rad redan utvecklade test. Om så är möjligt, bör man, redan innan den egentliga faktoranalysen genomförs, bestämma vissa statistiska kriterier, som t ex testens reliabilitet.

4. I många faktoranalyser görs den inskränkningen, att man enbart analyserar denna variabelgrupp (stickprovet). Man får på detta sätt visserligen upplysning om den interna faktorstrukturen för attributområdet i fråga. På vilket sätt de ur denna analys resulterande faktorerna skiljer sig från de faktorer, som man har funnit i andra områden, kan man dock ej avgöra.
5. För att undvika denna begränsning bör man utvidga variabelstickprovet med s k "markeringsvariabler" från redan kända faktorer, som kan vara av betydelse för det analyserade variabelområdet. För att man skall kunna välja lämpliga markeringsvariabler krävs det en ingående kännedom om den för undersökningen relevanta litteraturen. För varje sådan känd eller definierad faktor krävs det minst två markeringsvariabler för att kunna säkerställa, att dessa variabler dyker upp i de gemensamma faktorerna. Detta attributstickprov (variabelstickprovet + markeringsvariabler för de utvalda relateringsfaktorerna) undersöks och faktoranalyseras sedan på en och samma grupp av individer.
6. I faktoranalyser bör man principiellt undvika en flerfaldig utvärdering av test. Kommer t ex två variabler ur ett och samma test, ingår i dessa t ex samma observations- och mätfel, vilket har till följd att dessa variablers korrelation också omfattar felvariansen. Detta medför att modellen för flera gemensamma faktorer invalideras.
7. Även algebraiskt härledda variabler bör undvikas. I stället bör motsvarande råvärden användas. Summor eller differenser av testvärden medför principiellt inga nya faktorer. Dessa faktorer finns redan i de variabler ur vilka dessa beräknats. Det kan tilläggas, att införandet av härledda variabler bl a kan medföra tolkningssvårigheter.
8. Att faktoranalysera variabler, som bygger på ipsativa mätningar kan medföra svårigheter. En ipsativ mätning kan definieras på flera olika sätt. För definition, se Horst, 1963, ss 291-295.

Faktormodellen och de ur denna modell utvecklade metoderna för faktoranalys gäller uteslutande för kvantitativa variabler, som uppfyller kraven på en intervallskala. I vissa fall kan man dock också faktoranalysera s k kvalitativa data som åtminstone kan mätas i två klasser. Detta slags specialfall skall dock ej behandlas här.

4.2 Personstickprov

I allmänhet bestämmer redan frågeställningen personpopulationen. Utöver de allmänna riktlinjerna för urvalet av ett personstickprov bör man i en faktoranalytisk undersökning se till att stickprovet ej selegerar kraftigare med hänsyn till vissa enstaka variabler dvs att stickprovet ej är mera homogent i vissa variabler och mindre homogent i andra. Variabler, vilkas spridning är reducerad, korrelerar i regel lägre med de resterande variablerna. De förstnämnda variablernas kommunalitet underskattas och faktoranalysen ger i sådana fall bristfällig upplysning om variablernas sammansättning. Olämpliga för faktoranalysen är kluster, som är sammansatta ur icke representativa stickprov från olika populationer. I sådana fall finns det ej någon naturlig population, till vilken man kan generalisera korrelationerna och faktorerna. Därtill kommer att denna atypiska stickprovssammansättning medför atypiska attributkorrelationer samt resulterar i faktorer, som ej existerar i en naturlig population.

4.3 Stickprovets storlek

Metodexperiment visade att faktorladdningar med f ö lika betingelser är behäftade med väsentligt större stickprovsfel än korrelationskoefficienter. Faktoranalytiska undersökningar kräver därför större personstickprov än enkla korrelationsstatistiska utvärderingar.

Som ett absolut minimum måste man därför kräva en stickprovsstorlek av n 60-80. (Vissa äldre faktoranalyser genomfördes med n 30 eller mindre.)

I alla de fall, där det är möjligt, bör en faktoranalys dock baseras på n som är minst 100-150. Attributkorrelationer ligger till grund för faktoranalysen. Är dessa utsatta för stora stickprovsfel, representerar också faktorladdningar enbart mycket dåliga skattningar. Detta kan åskådliggöras med följande exempel: Medelfelet för en korrelation på $r = +0.80$ ligger för $n = 40$ omkring +0.59 och +0.91 och för $n = 150$ är medelfelet fortfarande mellan +0.71 och +0.86. Detta betyder att korrelationerna inte är statistiskt säkerställda ens med hänsyn till första decimalen och ett underlag av $n = 150$.

Men vid bestämning av personstickprovet måste storleken också sättas i relation till antalet variabler, som skall analyseras och därvid uppfylla vissa minimikrav. Helt allmänt gäller att antalet individer (n) skall vara större än antalet manifesta variabler (m), dvs $n > m$. Hur

7 1 0 4 2 2 0

mycket större n bör vara i förhållande till m finns det skilda uppfattningar om. Dessa beror bl a på om faktoranalysen är deskriptiv eller hypotesprövande (för beskrivning, se Pawlik, 1968, ss 85, 167).

5. NÄR SKALL MAN EJ FAKTORANALYSERA?

Begreppet faktoranalys har blivit ett honnörsord i vissa delar av den psykologiska och pedagogiska litteraturen. Ibland har man en känsla av att en viss författare med begreppet faktoranalys vill ge uttryck för uppfattningen att hans undersökning ligger utanför all kritik, eftersom han har använt sig av en faktoranalysteknik. Men eftersom det finns ett antal studier, där faktoranalysen använts på ett felaktigt sätt, kan det vara på sin plats, att nedan redogöra för en rad försiktighetsåtgärder, som utöver det som redan har sagts skall bidra till att faktoranalytiska metoder leder till adekvata lösningar. Redogörelsen följer Guilfords (1967, ss 309-317) argumentation. Först exempel på felaktiga användningar:

1. I vissa fall rör man sig med felaktigt uttagna variabler eller personer eller bådadera.
2. Ibland är det enda skälet för att man använder sig av en faktoranalytisk metod, att det finns interkorrelationer tillgängliga.
3. Den sämsta av alla tänkbara felanvändningar är, att man tror, att faktoranalysen kan förbättra en dåligt planerad undersökning, dvs man försöker "snygga upp" undersökningen genom att använda sig av en "fin" metod för analys av datamassan.

Om ett forskningsinstrument skall ge meningsfulla resultat, måste det användas på rätt plats och på rätt sätt. De vanligaste felanvändningarna - jämför exemplen ovan - kan ej ursäktas med att man ej är förtrogen med litteraturen om faktoranalysen eller ej haft tid att sätta sig in i denna litteratur. Många felanvändningar och missbruk skulle kunna ha undvikits, om försöksledaren eller experimentatorn hade iakttagit de vanliga reglerna för experimentell kontroll och för stickprovsurval.

Användningen av mera komplicerade statistiska metoder som faktoranalysen tillåter ej att man glömmer de vanliga säkerhetsåtgärder, som gäller för vetenskapliga, dvs systematiska, observationer.

Under olämpliga observationsförhållanden kan de erhållna data ge sken av en regelbundenhet i materialet, men detta är då felvisande, om ej fiktivt.

Det finns nämligen ej någon statistisk magi som ger en väl ordnad bild av datamassan när de analyserade data ej tillåter en meningsfull struktur. Diskussionen av olika möjliga felkällor sker i första hand med

hänsyn till sådana faktoranalyser som har rotation av referensaxlar som centralt moment.

5.1 Några allmänna fel

1. För många faktorer extraheras för det mesta med hänsyn till antalet experimentella variabler. Extraheras för många faktorer i förhållande till antalet variabler, utesluter detta en bra rotationslösning.

Thurstone har upprepade gånger poängterat att positionen för varje referensaxel i en rotationslösning måste överbestämmas. En bra regel är att man måste ha åtminstone tre variabler för varje faktor som väntas bli extraherad.

Det krävs ett antal olika test, som alla måste ha icke triviala laddningar på en faktor för att indikera faktorns allmängiltighet ("generality of the factor") och kunna definiera den på ett tillfredsställande sätt.

2. För många experimentella variabler är faktoriellt för komplexa. Rotationen och tolkningen skulle kunna vara mycket enklare, om varje variabel hade varit av komplexitetsgraden ett dvs att variabeln mäter enbart en allmän faktor på ett påtagligt sätt.

3. Ibland händer det att en väsentlig faktor ej extraheras på grund av att denna faktor enbart är representerad i en enda experimentell variabel.

För att en allmän faktor ("common factor") skall kunna extraheras med hänsyn till en bestämd experimentell variabeluppsättning krävs det, att fler än en variabel har en substantiell laddning i denna faktor. (Det antas här att man använder sig av vanlig praxis - att använda kommunaliteter i huvuddiagonalen av korrelationsmatrisen.) I annat fall förblir variabeln som specifik varians.

Detta faktum är i allmänhet ej till något större besvär eftersom det enbart betyder att man har misslyckats att extrahera en faktor, som finns i materialet och som i annat fall hade kunnat extraheras.

När man extraherar ett mindre antal faktorer än vad som finns i materialet, betraktas de övriga icke extraherade faktorernas varians som specifik varians.

4. Att ej extrahera ett tillräckligt antal faktorer betyder att man kanske missar empiriskt viktiga och intressanta dimensioner.

Erfarenheter visar att det är lönt att extrahera ett riktigt antal faktorer. De flesta kriterier på bestämningen av hur många faktorer som skall extraheras är ej baserade på en tillfredsställande teori.

Det kan ej medföra någon olägenhet att extrahera för många faktorer. Om det t ex extraheras fler faktorer än det mest gynnsamma antalet, så leder rotationen till ett förkastande av en eller flera av dessa faktorer, vilka då kvarstår som ej tolkbara. Att extrahera en eller två extra faktorer underlättar ofta tolkningen av den allmänna faktorstrukturen. Vissa experimentatorer syns vara rädda för att den sista faktorn ej representerar något annat än en "felvarians". Det syns som om de trodde att all varians vid extraheringen till en viss punkt är allmän-faktor-varians och att därefter all extraherad varians enbart består i felvarians. Det faktum att den senare extraherade variansen hjälper till att klargöra rotationsstrukturen och underlätta en psykologiskt meningsfylld tolkning av faktorerna är på sätt och vis ett bevis för att sann varians extraheras såväl i början som i slutet av extraheringsförfarandet.

Det bästa kriteriet vid extrahering av faktorer är storleken av de större faktorladdningarna. Så länge dessa är tillräckligt stora för att i rotationsprocessen kunna bidra något till andra faktorer, eller i den mån de passar in i något meningsfyllt sammanhang, så länge skall man fortsätta att extrahera faktorer.

5. Korrelationskoefficienter som ligger till grund för faktoranalyser kan vara oäkta.

Thurstone har ofta påpekat att korrelationer mellan variabler som användes i faktoranalyser, bör vara sådana att det ej torde finnas skäl för antagandet av en kovariation med undantag för en kovariation i de allmänna faktorerna.

Specifik varians och felvarians bidrar i vissa situationer, t ex vid användningen av personlighetsinventeringar, till korrelationer som egentligen ej är tillåtna. I en sådan situation tilldelar man ett enskilt item vikter för fler än en egenskapsvariabel.

Om ett antal items ingår i två olika summa-variabler uppkommer frågan, i vilken utsträckning korrelationerna av ett sådant multipelmarkerat ("multiple-scored") undersökningsinstrument verkligen motsvarar den relation bland variablerna, som testkonstruktören tror sig mäta. I vilken utsträckning dessa variabler representerar en tillfällig förbindelse av specifik varians och felvarians är okänt. Frågan, som man måste ställa, är, varför testvärdena på ett test korrelerar så kraftigt med varandra. Man måste då hålla isär två olika ting,

(a) å ena sidan faktorerna och

(b) å andra sidan värden, som var avsedda för att mäta dessa faktorer.

6. Ibland används ipsativa värden i en faktoranalys. Ipsativa mätningar kan bäst definieras som kontraster till normativa mätningar. En normativ mätning innebär att det för varje attribut finns en skala på vilken individpopulationen antar en viss fördelning. En ipsativ mätning innebär att det finns en skala för varje individ och populationen av individens attribut-värden fördelar sig på denna skala. Normativa värden används för att ange inter-individuella skillnader avseende ett attribut. Ipsativa värden används för att indikera intra-individuella skillnader i ett antal attribut.

Den vanliga faktoranalysmetoden för att analysera normativa värden är R-tekniken. För att analysera ipsativa värden är den lämpliga analystekniken en Q-faktoranalys (jfr kap. 3). Använder man sig dock av ipsativa värden, när man tillämpar R-tekniken eller normativa värden när man tillämpar Q-tekniken, bör man uppmärksamma speciella risker, som är förknippade med ett sådant förfaringssätt.

6. FAKTORANALYS KONTRA KOMPONENTANALYS

6.1 Fullständiga och reducerade faktor- eller komponentlösningar

Formellt sett hör faktoranalysen till den multivariata försöks- och analystekniken.

Lyckas man med att tillfredsställande förklara mätningarna i m variablerna $x_1, \ldots, x_m$ ur $k < m$ hypotetiska variabler, dvs att från dessa hypotetiska variabler på ett tillfredsställande sätt förklara de m manifesta variablerna, är faktoranalysen en ekonomisk och därmed vetenskapligt acceptabel metod för att presentera detta dataområde. Modellen är definierad som:

$$z_{ij} = \sum_{p=1}^{k} a_{ip} f_{pj} + q_{ij}$$

Målsättningen med alla faktoranalyser är alltså att på ett "enklare" sätt, dvs med ett mindre antal av varandra oberoende och gemensamma faktorer, kunna förklara ett visst egenskapsområde.

Mycket svårare är det däremot att uppfylla tilläggskravet, att även relationerna mellan faktorerna och de observerade egenskaperna skall

vara av enklast möjliga struktur.

Resultatet av en faktoranalys kallas en fullständig faktorlösning, när man ur denna lösning kan förklara hela variansen för de m experimentella variablerna, dvs om de experimentella observationerna har reproducerats exakt. För att detta skall gälla, måste i allmänhet antalet faktorer vara lika med antalet variabler. För g-faktormodellen och för modellen med flera gemensamma faktorer, måste det t o m vara större. Betecknas det fullständiga antalet faktorer med r väntas i allmänhet $r \geqslant m$.

De r faktorer som ingår i en fullständig faktorlösning är såväl innehållsmässigt som statistiskt av olika betydelse. Medan ett litet antal faktorer svarar för den största delen av variansen finns det också många "mindre faktorer", som i mycket liten utsträckning bidrar till förklaringar av den allmänna attributvariansen. Dessutom finns det ett antal specifika faktorer som ej får tillmätas en mera allmän betydelse.

Vid resultattolkningen kan dock såväl de mindre som de specifika faktorerna vara till hjälp, men dessa utesluts vanligen ur redovisningen.

Denna förenklade presentation av m stycken variabler med hänsyn till k stycken faktorer, som upptar den största delen av variansen av r faktorerna, kallas en reducerad faktorlösning. Denna faktorlösning innehåller då enbart de faktorer som anses vara tolkbara. I detta fall är då inte bara $k < r$ utan alltid också $k < m$.

Till skillnad från den fullständiga faktorlösningen tillåter k faktorerna i den reducerade faktorlösningen endast en approximativ bestämning av experiment- eller försöksresultat.

Varje faktorproblem är intimt förknippat med besvarande av frågan: På vilket sätt bestämmer man i praktiken när ett tillägg av en ny faktor enbart medför en försumbar reduktion av restvariansen? Detta problem är hittills ej matematiskt löst på ett tillfredsställande sätt, men det finns en rad empiriska metoder för att lösa problemet (se Pawlik, 1968, ss 167-172).

En fullständig komponentanalys innebär, att de m stycken experimentella variablerna omstruktureras till en ny uppsättning r stycken variabler. Dessa r stycken nya variabler är oberoende av varandra och svarar tillsammans för hela attributvariansen. I regel är r = m. Dessa nya variabler är lineära kombinationer av de m experimentella variablerna.

Den reducerade komponentanalysen avser, i likhet med den reducerade faktoranalysen, ett mindre antal komponenter än r. Detta mindre antal k svarar för den största delen av den totala variansen. Enligt

definition nedan har komponentanalysen speciella variansegenskaper. Det finns nämligen ingen annan uppsättning av k eller färre komponenter som förklarar en större del av den totala variansen.

6. 2 Komponentanalysmodellen

När man faktoranalyserar den totala summa-variansen, har man gett upp den grundläggande faktormodellen och utför i stället en komponentanalys. Begreppet komponent användes av Hotelling ("principal component") för att undvika en förväxling med det matematiska faktorbegreppet (med innebörden: faktor i en produkt). Ur teoretiska och även ur praktiska synpunkter kan denna teknik vara otillfredsställande, men för det mesta skiljer den sig i praktiken bara i mycket ringa grad från faktoranalys-resultatet. Modellen är definierad som

$$z_{ij} = \sum_{p=1}^{k} a_{ip} G_{pj}$$

där:

z_{ij} = standardiserade observationsvärden

a_{ip} : komponentladdningar och

G_{pj} : standardiserade och okorrelerade komponentvärden.

Två egenskaper skiljer denna modell från de andra mera speciella faktormodellerna:

1. Det fullständiga antalet komponenter är lika med antalet experimentella variabler.
2. Det görs ingen skillnad mellan gemensamma komponenter och specifika komponenter. I analysen ingår den totala variansen.

Skälet för att man utför en komponentanalys är följande: Genom en matematisk modell transformerar man den ursprungliga variabeluppsättningen till en ny exakt noll-korrelerad variabeluppsättning, dvs till komponenter.

Den första av dessa komponenter tar upp ett maximum av variansen av de ursprungliga variablerna. Den andra komponenten tar upp ett maximum av variansen i residualmatrisen, sedan den första komponenten har extraherats, etc. Summan av alla komponenterna är lika med den totala variansen.

Vanligen söker man ett fåtal komponenter för att ge en bra approximering av den ursprungliga datauppsättningen. Dessa fåtaliga komponenter visar vanligen en bra överensstämmelse med vanliga faktoranalysmetoder, dvs man får resultat som är mycket lika varandra.

Den mest direkta användningen av komponentanalysen medför enbart en omstrukturering av variablerna till en ny uppsättning av m stycken okorrelerade variabler med fallande varians och har knappast någon direkt dimensionsanalytisk anknytning. Den reducerade komponentanalysen innebär däremot en analys av dimensioner i materialet (Mårdberg, 1969, s 52).

Vanligen föredrar man en faktoranalys för att den ger en mera ekonomisk beskrivning av variabeluppsättningen. Men huvudsakligen använder man faktoranalysmetoden oftare eftersom modellen medger en isolering av felfria och icke specifika faktorer för att representera korrelationerna.

Den totala variansen spaltas sålunda upp i olika delar. I tabell 2 ges en sammanfattning av skillnaderna mellan komponent- och faktoranalys.

Tabell 2. Faktoranalys kontra komponentanalys

Faktoranalys	Komponentanalys
Dimensionsanalys	Dimensionsanalys
Statistisk modell	Matematisk modell
Lineära kombinationer av de manifesta variablerna med hänsyn till den gemensamma variansdelen.	Lineära kombinationer av de manifesta variablerna med hänsyn till den totala variansen
Individernas faktorvärden kan endast skattas	Individernas komponentvärden kan exakt bestämmas
Modellen förklarar kovariansen	Modellen förklarar den totala variansen
Faktorlösningen sker iterativt, modellens parameter estimeras, där $r > m$	Komponentlösningen sker iterativt. Komponenter får man genom en direkt omstrukturering till en ny, okorrelerad variabeluppsättning, där $r = m$
Rotering - faktorladdningsmatrisen består i en estimering av de manifesta variablernas lägen i de latenta variablerna	Ingen rotering - komponentmatrisen består i en direkt vägning av de manifesta variablerna. Komponenterna tar successivt upp den största möjliga variansen
Modellen ger en isolering av de felfria och icke specifika faktorerna	Modellen differentierar ej mellan felfria och icke specifika komponenter
Faktoranalys är en ekonomisk beskrivning av materialet	Komponentanalysen är endast som reducerad komponentanalys en ekonomisk beskrivning av materialet

7. ATTRIBUTVARIANS

7.1 Uppspaltning av attributvariansen

Ett centralt begrepp i samband med olika statistiska analystekniker är en variabels varians och vilka olika variansdelar som är aktuella i samband med diskussion av vilka variansdelar olika statistiska metoder analyserar. En experimentell variabels x totala varians består av en systematisk variabel t_i och av en felvariabel e_i, dvs

$$x_i = t_i + e_i \qquad (8)$$

Grundläggande för skillnaden mellan komponentanalys och faktoranalys är sättet att behandla en variabels eller variabeluppsättnings totala varians. Medan i komponentanalysen ej tas hänsyn till olika variansdelar, specificeras olika delar i faktoranalysmodellen. De variansdelar för en variabel, som kan generaliseras till datamatrisen är:

1. gemensam varians som betecknas h_i^2 och är den del av variansen som förklaras av de k gemensamma faktorerna,
2. restvarians u_i^2 som är den del av variansen som ej kan förklaras av k faktorerna.

För standardiserade variabler medför punkt 1 och 2 att

$$h_i^2 + u_i^2 = 1 \qquad (9)$$

I en fullständig faktorlösning är k = r och restvariansen $u_i^2 = 0$. Därav följer att $h_i^2 = s_i^2 = 1$ (10)

där

$$h_i^2 = \sum_{p=1}^{k} a_{ip}^2 \text{ och} \qquad (11)$$

$$s_i^2 \sum_{p=1}^{k} a_{ip}^2 + u_{ip}^2 = 1 \qquad (12)$$

r faktorerna förklarar alltså totalsumman av attributvariansen. Geometriskt kan den positiva kvadratroten ur kommunaliteten för en variabel förklaras som vektorns längd i den reducerade faktorrymden.

I samband med uppspaltningen av variansen i olika delar förekommer även begrepp som

3. unik varians. En variabels unika varians är variabelns totala varians minus den gemensamma variansen. Begreppet är alltså identiskt med restvariansen. Denna variansdel kan ej prediceras med utgångspunkt från andra variabler i datamatrisen. Variansdelen brukar man ofta dela upp i specifik varians och felvarians.

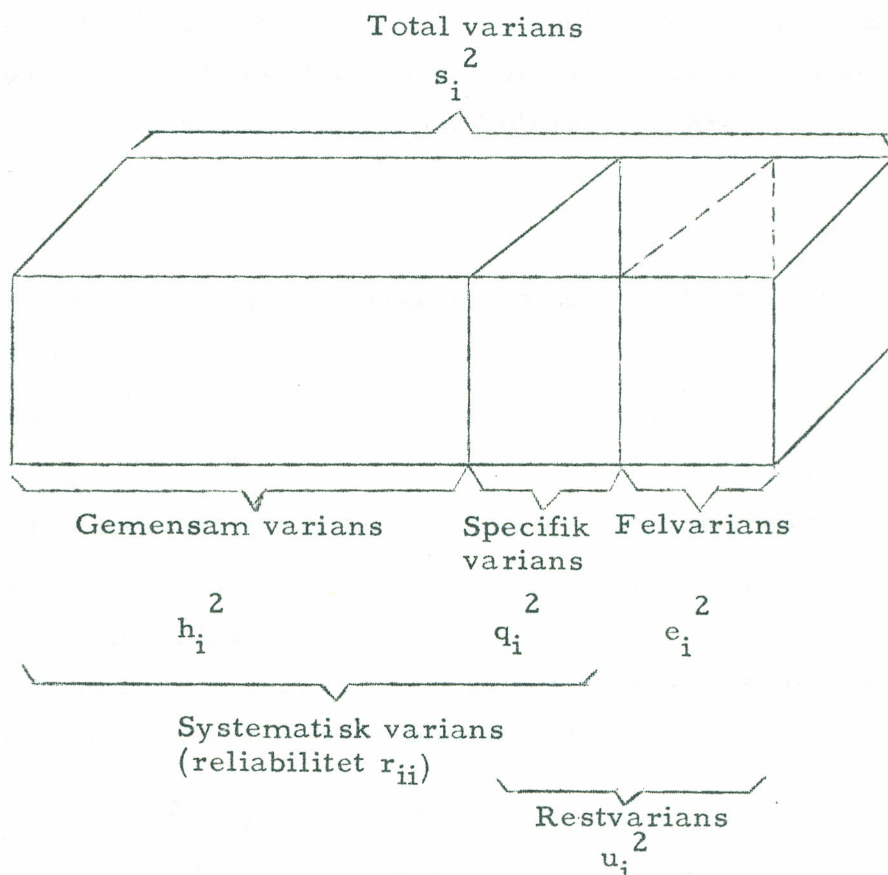
7104228

4. Specifik varians är den del av restvariansen som icke är slumpmässig. Denna variansdel hänför sig till en variabels specificitet, dvs den variansandel som variabeln har oberoende av alla andra variabler i matrisen.
5. Felvarians är den del av restvariansen som är slumpmässig, osystematisk.
6. Systematisk varians är den del av den totala variansen som defineras av den gemensamma och den specifika variansen. Den systematiska variansen är enligt definitionen variabelns reliabilitet.

Därav följer att man får följande ekvivalenta ekvationer

$$s_i^2 = \left\{ \begin{array}{l} h_i^2 + \overbrace{u_i^2}^{} \\ \underbrace{h_i^2 + q^2}_{} + e_i^2 \\ r_{ii} + e_i^2 \end{array} \right\} = 1 \qquad (13)$$

Med utgångspunkt från dessa ekvationer presenteras relationen mellan dessa olika variansdelar grafiskt i figur 5.



Figur 5. Relation mellan variansdelarna.

Om man antar att för varje variabel finns givna h_i^2 och u_i^2, blir den totala variansen för en variabel:

$$\text{total varians} = \sum_{i=1}^{m} s_i^2 = m. \qquad (14)$$

7104229

Den kvadrerade multipla korrelationen för ett kriterium med en prediktor får man genom att man parvis multiplicerar regressionskoefficienten med motsvarande prediktor-kriterium-korrelation och adderar dessa, dvs

$$R_{ik}^2 = \sum_{p=1}^{k} r_{ip} \beta_{ip} \qquad (15)$$

I faktoranalysen är laddningen a_{ip} samtidigt regressionskoefficient för den stanardiserade variabeln z_i när faktorerna är standardiserade och oberoende av varandra och dess faktorkorrelationer. Summan

$$\sum_{p=1}^{k} a_{ip} \cdot a_{ip} = \sum_{p=1}^{k} a_{ip}^2 \qquad (16)$$

måste därför vara identisk med den kvadrerade multipla korrelationen R_{ik} mellan z_i och k faktorerna, dvs

$$R_{ik}^2 = \sum_{p=1}^{k} a_{ip}^2 = h_i^2 \qquad (17)$$

Kommunaliteten för en variabel är alltså variabelns kvadrerade multipla korrelation med k faktorerna. Som variansbidrag av den p:te faktorn betecknas summan av de kvadrerade faktorladdningarna för den p:te kolumnen i faktormatrisen, dvs variansbidrag av f_r är

$$V_p = \sum_{i=1}^{m} a_{ip}^2$$

Summan av alla k faktorernas variansbidrag kallas det totala variansbidraget. Det är

$$\sum_{p=1}^{k} V_p = \sum_{p=1}^{k} \sum_{i=1}^{m} a_{ip}^2 \qquad (18)$$

dvs den totala extraherade variansen, som är identisk med $\sum_i h_i^2$

Summan av restvarianserna u_i^2 kan kallas den totala restvariansen (felvarians och specifik varians)

Analogt med en enskild variabel gäller samma resonemang för hela datamatrisen, dvs för alla m variabler har vi

$$\underbrace{\sum_i s_i^2 = m}_{\text{total varians}} = \underbrace{\sum_{p=1}^{k} V_p}_{\text{totalt variansbidrag av k faktorer}} + \underbrace{\sum_{i=1} u_i^2}_{\text{total restvarians}} \qquad (19)$$

7. 2 Kommunalitetsproblemet

Guttman har visat att R_i^2, den multipla korrelationen i kvadrat mellan x_i och de övriga m-1 variablerna, är en nedre gräns för h_i^2. (Se Rozeboom, 1966, s 261!) Samtidigt gäller att kommunaliteten inte kan överstiga reliabiliteten. Tillsammans ger detta

$$R_i^2 \leq h_i^2 \leq r_{ii} \qquad (20)$$

Både r_{ii} och i synnerhet R_i^2 har använts som skattning av h_i^2, där den senare skattningen blir bättre ju större m är, förutsatt att personstickprovet är stort.

I faktoranalysen arbetar man ej med kommunaliteter som ligger över variabelns reliabilitet. En konsekvens är, att om vi minskar variabelns egentliga reliabilitet, får vi som resultat en minskning i faktorernas laddningar.

I faktoranalysen är kommunalitetsproblemet av central betydelse, eftersom kommunalitetsskattningen är avgörande för (1) hur mycket av variansen som extraheras och (2) för bestämningen av de specifika variablerna i materialet.

Som tidigare nämnts är syftet med modellen att representera m variablers korrelationer genom k gemensamma faktorer, där $k < m$. Variansandelen i de gemensamma faktorerna kan dock endast bli 1.0 om all varians för varje enskild variabel tas upp genom de gemensamma faktorerna. Men att arbeta med gemensamma faktorer som tar upp <u>all</u> varians är av två skäl mindre önskvärt:

1. Vi antar att de gemensamma faktorerna är fria från fel-variationer. Detta antagande stödjes genom att faktorerna är latenta variabler och fel introduceras enbart genom själva mätningen i de manifesta variablerna. Eftersom alltså en del av variansen i variablerna uppstår genom mätningsfel kan vi ej förvänta oss att faktorerna tar upp denna del av variansen.

2. Det andra argumentet implicerar att kommunaliteten borde vara mindre än reliabiliteten. Detta antagande baseras på att vi förväntar oss en del av den felfria variansen som specifik för en given variabel. Vilken storleksordning denna del av variansen har, är beroende på variablernas specificitet. Den totala variabeluppsättningen definierar alltså de gemensamma faktorernas kommunalitet.

Om man antar att $h^2 = 1.0$, faktoriserar vi både den gemensamma och den specifika variansen. Resultatet är: för många faktorer. Om vi

i stället använder oss av en viss reliabilitetsskattning för h^2 får vi fortfarande för många faktorer.

Om vi använder oss av en underskattning av kommunaliteten resulterar detta i imaginära faktorladdningar.

8. EXTRAHERING AV FAKTORER OCH EXTRAHERINGSKRITERIER

Det finns ett antal olika metoder för att extrahera faktorer dels successivt, dels simultant.

En successiv extraheringsmetod är t ex att man vid varje extraheringstillfälle bestämmer en faktor. Då den första faktorn extraherats, bestämmes i detta förfaringssätt en residualmatris. Från denna residualmatris extraheras sedan nästa faktor osv.

En annan metod, som inte är så vanlig - men i vissa fall erbjuder fördelar - är att man även här successivt extraherar varje faktor. Vid användningen av denna metod behöver man emellertid inte bestämma en ny residualmatris för varje gång.

Successiva extraheringsmetoder innebär att en bestämd faktor extraheras och på så sätt tages bort från korrelationsmatrisen. Detta är den mest väsentliga skillnaden från de multipla grupp-faktoriseringsmetoderna. Dessa innebär, att alla faktorerna extraheras simultant, samt att det krävs en hypotes om hur många faktorer, som är nödvändiga för att på ett tillfredsställande sätt approximera korrelationsmatrisen.

Maximum varians-faktorer avser sådana extraheringar, som medför att den första faktorn upptar maximal varians för en viss variabeluppsättning, som kan representeras genom en enda faktor. Nästa faktor, som extraheras, tar upp näst största delen av variansen, som ej kunde tas upp av den första faktorn. Denna lösning stämmer dock enbart om iterationsförfarandet har genomförts tillräckligt, dvs med önskad, numerisk noggrannhet.

Hotelling utvecklade den numera mest använda extraheringsmetoden, den iterativa principalaxelmetoden. Den kan i korthet beskrivas på följande sätt:

Denna metod är en successiv extraheringsmetod. Först skattas laddningarna av variablerna i den <u>första</u> principalaxelfaktorn, därefter laddningarna i den andra faktorn osv, tills man har fått det önskade antalet k, resp tills alla m faktorerna är extraherade.

Detta förfaringssätt är dessutom <u>iterativt</u>, dvs variablernas laddningar i f_p beräknas ej direkt, utan man utgår från s k "provvärden",

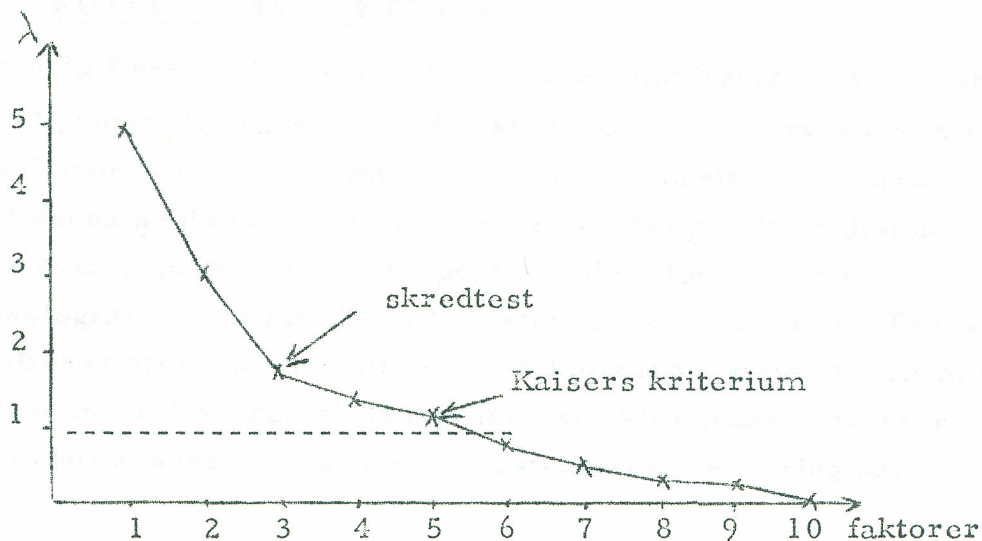
som sedan förbättras genom successiva approximationer (iterationer), tills man har uppnått den önskade noggrannheten.

Faktormatrisen ligger till grund för faktortolkningen. Men faktoranalysen ger än så länge enbart basen för faktorrymden, dvs r ömsesidigt ortogonala vektorer som representerar rymden.

Ju närmare approximationsvektorn från början ligger den korrekta vektorn, desto mindre antal iterationer behövs i regel för att lösa faktorvektorn. Efter beräkningen av en given faktorladdningsvektor måste man avgöra, om ett tillräckligt stort antal faktorer extraherats.

Det finns flera mer eller mindre strikta kriterier för extrahering av ett "tillräckligt" antal faktorer. Eftersom varje successiv faktor i principalaxelmetoden ej tar upp mer (och vanligen mindre) varians än den tidigare faktorn kan man fortsätta med att extrahera faktorer med egenvärden som är $\lambda = \geq a$. Alla andra faktorer kan anses sakna betydelse. Man brukar sätta a = 1. Metoden sägs vara speciellt lämplig vid utnyttjandet av principalaxelmetoden. Detta kriterium har bl a betecknats som "Kaisers kriterium". Man kallar det även Guttmans svagaste kriterium för antalet signifikanta faktorer (Guttman, 1954, ss 149-161). Cattell anser att Kaisers kriterium troligen är mest tillförlitligt när antalet variabler ligger mellan 20 och 50. Är antalet variabler mindre än 20 finns det en viss risk för en konservativ extrahering av antalet faktorer. Ingår däremot fler än 50 variabler i analysen extraheras för många faktorer vid användningen av Kaisers kriterium (Child, 1970, ss 43-44).

Ett annat kriterium för bestämningen av antalet faktorer, som skall extraheras är att man undersöker om det i fördelningen av faktorernas latenta rötter (λ) finns ett distinkt brott. Kriteriet kallas enligt Cattell (1963) "scree test" (skredtest). Vid användningen av detta kriterium utnyttjar man en grafisk presentation av λ-fördelningen (figur 6) för att bestämma "cut off"-punkten i kurvan.



Figur 6. Kaisers kriterium och skredtest.

Man börjar med det största egenvärdet. Först faller kurvan snabbt, men övergår vid tredje faktorn (cut off) i en lineär relation. Använder man sig av Kaisers kriterium skulle analysen ha resulterat i en extrahering av 5 faktorer. Användningen av Cattells skredtest skulle däremot ha lett till en extrahering av tre faktorer. Men skredtestet kan dock också medföra att man extraherar flera faktorer än vid användningen av Kaisers kriterium, dvs brottet i kurvan uppstår först efter $a \leq 1$. Använder man sig av skredtestet krävs det att man utför två faktoranalyser, en för att bestämma kriteriet och en för att extrahera faktorerna.

Ytterligare en tumregel kan vara att man bestämmer att antalet faktorer som extraheras förklarar 80 % av den kumulativa variansen, dvs k faktorer väljer man så att $\sum_{p=1}^{k} p_j \approx 0.80$ (Mårdberg, 1969, s 58).

Morrison (1967, s 228) anser att k = 4 eller 5 borde svara för 75 % av den kumulativa variansen. Att extrahera flera faktorer är ofta av ringa intresse anser Morrison, eftersom man ändå får tolkningssvårigheter.

Det är också möjligt att använda sig av signifikanstest för att bestämma antalet faktorer som skall extraheras.

Signifikanstest i samband med faktoranalysen tar i regel hänsyn till en speciell faktormodells egenskaper och är därför inte utan vidare utbytbara mellan olika modeller.

För r faktorer testas om dessa kan reproducera korrelationsmatrisen R tillräckligt bra.

Man kan successivt testa för r = 1, 2, ... k faktorer. Dessa successiva χ^2-test blir dock ej oberoende av varandra (se Morrison, 1967, s 275, Pawlik, 1968, ss 87-88, 167-173). Det r-värde som ger minsta χ^2-värde kan betraktas som den "bästa" lösningen.

9. ROTERING AV FAKTORER

I princip finns det ett oändligt antal olika möjligheter att extrahera faktorer. Först genom tilläggskraven på faktorerna, som uttrycks med hjälp av rotationsmetoder, bestämmer man en viss position för faktorerna. Men denna bestämning är subjektiv med hänsyn till modellen. Det bör dock även påpekas, att denna position dessutom inte alls behöver vara psykologiskt meningsfull. Alla metoder som baseras på Thurstones rotationskriterium, dvs att rotera faktorerna i syfte att allt närmare approximera faktorernas tänkta idealstruktur (enkel-struktur-kriteriet) kräver iterativa rotationstekniker. Rotering kan

göras antingen rent grafiskt eller analytiskt. Varimax-kriteriet är ett exempel på en analytisk rotationsmetod.

Åtminstone två krav bör vara uppfyllda för en analytisk rotationsmetod (samt för vilken annan rotationsmetod som helst):

1. faktorladdningarna bör vara relativt invarianta med hänsyn till den persongrupp, från vilken data har insamlats,
2. en delgrupp av faktorladdningarna bör vara relativt invariant, oberoende av det speciella testbatteri (den variabeluppsättning), som innehåller denna delgrupp.

Den största svårigheten vid användningen av analytiska metoder är att de i stor utsträckning påverkas av det speciella variabelurvalet som sedan ingår i korrelationsmatrisen (Horst, 1965, s 419).

Kaiser (1958, 1959) utvecklade varimax-kriteriet som är ett ortogonalt och analytiskt rotationsförfarande. Kriteriet härleds ur det s k Q_3-kriteriet. Q_3-kriteriet innebär att variansen för de kvadrerade faktorladdningarna skall bli maximal. Varimax-kriteriet baseras, till skillnad från Q_3, där man använder totalmedelvärde, på resp.kolumnvärde av de kvadrerade faktorladdningarna.

Variansanalytiskt uttryckt betyder detta: I stället för att använda sig av den totala variansen av de kvadrerade faktorladdningarna använder man vid tillämpningen av varimax-kriteriet variansen inom kolumnerna av den roterade faktormatrisen (V_p) som är resultatet av den oroterade faktormatrisen (A_p).

Detta kriterium har dock samma nackdel som kvartimax-kriteriet: för det mesta blir en faktor en generell faktor, dvs har så gott som inga nolladdningar. Detta beror på det faktum, att rotationen i allt för hög grad blir bestämd av variablernas kommunalitet. För att undvika detta fenomen rekommenderade Kaiser, att man vid beräkningen av rotationsvinkeln skulle normera samtliga attributvektorer.

Varimax-rotationen löses iterativt. Varje iteration består av enskilda cykler, dvs för varje möjlig kombination av två kolumner i faktormatrisen utförs en rotation. För den resulterande matrisen V_k beräknas sedan ett kriterie-värde och därvid utförs så många iterationer att kriteriet konvergerar till önskad noggrannhet. Ett exempel:

Tabell 3. Iterationsprocess

	Varimax-kriteriet
oroterade matris A_k	+.23048
efter 1. iteration	+.45050
efter 2. iteration	+.46143
efter 3. iteration	+.46149
efter 4. iteration	+.46149

Kaisers metod är känslig för avrundningsfel. Det stora antal beräkningar, som denna metod innebär, medför en ackumulering av avsevärda decimalfel, när antalet variabler är stort, t ex 500 variabler och ca 20-30 faktorer.

I praktiken har varimax-metoden visat sig vara ett pålitligt förfaringssätt för en första approximation av enkelstrukturen. Analytiskt roterade faktorlösningar kan man nästan undantagslöst förbättra med några ytterligare visuella rotationer för att hitta den bästa möjliga enkelstrukturen.

Varimaxmetoden tillhör en klass av metoder som har kallats blindroteringsmetoder. Deras gemensamma drag är att kriteriet för metodernas genomförande är internt, som t ex maximal varians i visst avseende. I motsats härtill finns s k målroteringsmetoder som förutsätter ett externt kriterium. Vid användningen ställer man en hypotes om hur många faktorer det bör finnas och om storleken på faktorladdningarna. Målroteringsmetoden ger en lösning som i någon mening approximerar den hypotetiska faktormatrisen så bra som möjligt. Är approximeringen "tillräckligt" god kan hypotesen anses rimlig. Metoder och program för denna typ av faktoranalys är tyvärr ännu ej tillfredsställande utvecklade.

10. IDENTIFIKATION OCH TOLKNING AV FAKTORER

10.1 Den geometriska tolkningen av faktorladdningar och faktorvärden: faktorrymden

En datamatris av typen individ gånger variabler, kan man geometriskt tolka på två olika sätt:

1. Variablerna uppfattas som koordinataxlar och
2. Fpp uppfattas som koordinataxlar.

Uppfattar man variablerna som koordinataxlar, får man en presentation av försökspersonerna som punktsvärm i en variabelrymd R_v.

Uppfattar man däremot försökspersonerna som koordinataxlar, får man en presentation av variablerna som en punktsvärm i en personrymd P_p. Den s k faktorrymden R_f får man genom att man låter faktorerna bilda koordinataxlarna. Denna faktorrymd är en analogi till variabelrymden. Den fullständiga faktorrymden R_f bildas alltså av r lineära och av varandra oberoende vektorer f, ... f_r. Eftersom dessutom dessa vektorer är ömsesidigt ortogonala bildar r faktorerna ett rätvinkligt koordinatsystem i faktorrymden.

För den reducerade faktorlösningen gäller principiellt samma resonemang. Här konstitueras av de första k av r faktorerna en k-dimensionell delrymd. Liksom variabelrymden är också i faktorrymden varje försöksperson representerad med en punkt. Koordinaterna för denna punkt bildas av faktorvärdena.

Principiellt kan man faktorisera varje slag av datamatriser, t ex råvärdematriser eller standardvärdematriser. Denna ordningsföljd innebär naturligtvis en successiv informationsförlust. Men endast i de fall där alla variabler presenteras i samma mätningsform, dvs där medelvärdes- och spridningsskillnader mellan variablerna kan tolkas, är det lämpligt att utföra en faktoranalys av råvärden.

10. 2 Faktorladdningarnas egenskaper

Antalet k faktorer bestäms på ett sådant sätt, att den genom dessa faktorer icke förklarade restvariansen av variablerna är så liten som möjligt.

1. Faktorladdningarna är de standardiserade regressions-koefficienterna för variablerna. Faktorladdningarnas definitionsekvation är alltså motsvarande regressionsekvation.

   En annan egenskap hos faktorladdningar - som alltid gäller när ortogonalitetskravet är uppfyllt - är följande: De observerade attributen och faktorerna är två alternativa system med hänsyn till de variabler, med vilkas hjälp försökspersonerna representeras och därför måste dessa korrelera med varandra. Faktorladdningarna ger uttryck för korrelationsgraden mellan de observerade attributen och faktorerna.

2. När ortogonalitet råder, är variablernas faktorladdningar lika med variablernas korrelation med faktorerna. Som korrelationskoefficienter kan därför faktorladdningarna (i ortogonalt fall) endast variera mellan -1, 0 och +1, 0.

Medan under punkt 1 faktorladdningarna förklaras statistiskt, tillåter punkt 2 en mycket enkel och åskådlig tolkning av faktorladdningar. Men i båda fallen baseras förklaringen till uppspaltningen av variansen på faktoranalysmodellen.

Det finns tyvärr än så länge ingen statistiskt tillfredsställande testmetod för att bestämma vilka laddningar, som kan anses vara nolladdningar och vilka laddningar, som måste anses vara substantiella laddningar. Tills vidare är vi därför hänvisade till empiriska värden. För det mesta använder man sig enligt Pawlik (1968, ss 182-183, 265-266) av följande tumregel:

1. Faktorladdningar mellan -.10 och +.10 betraktas som nolladdningar.
2. Faktorladdningar mellan -.20 och +.20 betraktas som tillfälliga avvikelser (slumpvariationer) från 0,00.

   Frågan i vilken utsträckning laddningar får avvika från noll och ändå fortfarande kunna betraktas som noll, får ej förväxlas med frågan, i vilken utsträckning en laddning måste avvika från noll för att den skall kunna vara interpreterbar. Vanligen tillämpar man i det andra fallet en mera konservativ gräns. Vad som är höga laddningar får man dock avgöra själv. Morrison (1967, s 290) visar t ex att medelfelet varierar mellan $0.7/\sqrt{N}$ till $1.6/\sqrt{N}$. Kanske skulle man, konservativt sett, kunna säga att laddningen för en viss variabel borde vara sådan att $a_{ip} \geqslant 4/\sqrt{N}$. Men denna problematik är icke utredd.
3. Faktorladdningar på $\pm$ .30 betraktas som signifikanta laddningar. Thurstone, Cattell och de flesta andra faktoranalytiker kräver att rotationen skall utföras som blind-rotation, dvs utan kännedom om den psykologiska betydelsen av variablerna. Bara på ett sådant sätt kan man garantera, anser dessa författare, att faktorlösningen inte enbart roteras i riktning mot förutfattade hypoteser, vilket skulle vara av ringa värde för en hypotesprövning. I större analyser hittar man dessutom ofta flera lika "meningsfulla" positioner för en faktor. Thurstone's enkelstruktur-kriterium används i dag i de allra flesta psykologiska faktoranalyser.

10.3 Identifikationsproblem

I regel kan man ställa upp flera tolkningshypoteser för en faktor. På grundval av detta faktum kan man därför aldrig avgöra om man kan likställa två faktorer. Alla faktoridentifikationsmetoder baseras på

parvisa jämförelser av faktoranalyser. Dessa parvisa jämförelser förutsätter, att de med varandra jämförda analyserna grundar sig på en gemensam bas, dvs analysen baseras antingen

1. på samma urval (stickprov) försökspersoner, eller
2. på åtminstone delvis samma variabeluppsättning.

Fall 1. (Bas: gemensamma fpp):
två analyser baseras på ett och samma personstickprov, men analyserna hänför sig till olika variabler.

Fall 2. (Bas: gemensamma variabler)
Två analyser utförs på två olika personstickprov men innehåller åtminstone delvis identiska variabler.

Faktoridentifikationsmetoden är olika för varje enskilt fall. Dessutom är det omöjligt att åstadkomma en direkt faktoridentifikation, när två faktoranalyser grundar sig samtidigt på olika försökspersoner och olika variabeluppsättningar. I sådana fall måste en ny undersökning genomföras, där man analyserar båda egenskapsstickproverna tillsammans.

10.4 Faktortolkning

Att tolka en faktor betyder att man försöker ställa upp hypoteser om faktorns psykologiska natur och betydelse. Dessa hypoteser bör man formulera på ett sådant sätt, att hypotesernas giltighet kan avgöras med hjälp av efterföljande experiment. Man borde enbart tolka sådana faktorer, som kan identifieras tillräckligt entydigt i minst två av varandra oberoende faktoranalyser, vilket då medför att de kan anses vara någorlunda säkra faktorer. Faktortolkningen utförs på den roterade faktorladdningsmatrisen. Ur variablerna utväljer man för varje faktor två olika grupper:

1. s k markeringsvariabler för faktorn i fråga, dvs variabler med absolut höga laddningar (ungefär från och med 0.50-0.60) och
2. s k hyperplanvariabler, dvs variabler med laddningar av storleksordningen mellan -.10 +.10.

Visar det sig, att markeringsvariablerna i en bestämd faktor i de resterande gemensamma faktorerna är representerade med låga laddningar, hänförs dessa variablers gemensamma varians huvudsakligen (om inte uteslutande) till denna faktor. Faktorn representerar då en beteendekomponent, som får sitt uttryck i dessa variabler.

Vid tolkningen av en faktor försöker man nu ange sådana psykologiska egenskaper, som är gemensamma för alla markeringsvariabler i respektive faktor och som samtidigt ej tillhör de variabler, som ligger

i hyperplanet. Hyperplansvariablerna ger alltså samtidigt bakgrunden, mot vilken man kontrasterar det gemensamma i faktorns markeringsvariabler.

För det mesta får man dock flera alternativa tolkningshypoteser. Vilken av dessa hypoteser, som kan anses vara mest rimlig, kan man dock enbart avgöra i en fortsatt försöksverksamhet. Vid en faktoranalytisk tolkning. måste man vara särskilt uppmärksam på flera olika felkällor. Några mera väsentliga exempel är följande:

1. specifika faktorer
   a) Har man extraherat för många faktorer, finns det för vissa faktorer endast en enda variabel med en substantiell laddning.
   b) Sådana faktorer bör man alltså ej tolka på samma sätt som gemensamma faktorer.
2. Specifika parfaktorer
   En faktor kallas parfaktor, då enbart två variabler laddar substantiellt.
   a) Är dessa två variabler lineärt beroende eller varandra tämligen lika (kanske t o m ur samma test eller ett paralleltest), representerar parfaktorn för det mesta en specifik varianskomponent med hänsyn till dessa båda variabler (ur ett bestämt test).
   b) Sådana parfaktorer bör ej tolkas som gemensamma faktorer och roteras ortogonalt i förhållande till alla andra faktorer.
3. Metodfaktorer
   a) Med metodfaktorer avses sådana faktorer, som ej baseras på testets innehåll, utan baseras på metodiska likheter i markeringsvariablerna, dvs i testformatet, tekniken för testadministrering etc.
   b) Det är mycket väsentligt att man uppmärksammar sådana gemensamma faktorer, som är metodfaktorer, för att undvika allvarliga feltolkningar.

En alldeles speciell noggrannhet och omsorg bör man tillämpa vid beteckningen av faktorerna för att undvika missförstånd. Har man i en hel serie av undersökningar kunnat säkra en bestämd hypotes, betraktas denna faktor preliminärt som säkrad.

För att underlätta kommunicerbarheten använder man sig i sådana fall av en kort beteckning, som innehåller faktorns mest väsentliga karakteristika, dvs man ger faktorn ett namn, t ex induktivt tänkande, emotionell balans etc.

Sådana faktorbeteckningar skall man hålla deskriptivt. Med andra ord, man bör ej blanda in oprövade hypoteser och teorier.

Beteckningar, som redan är teoretiskt förankrade i annat sammanhang, bör man absolut undvika, eftersom det lätt leder till missuppfattningar och otillåtna utvidgningar av en viss faktor.

10.5 Faktorladdnings- och faktorvärdematris

I allmänhet anses faktorladdnings-matrisen som den mest väsentliga matrisen. Vid en enkelstrukturrotering eller någon annan transformeringsteknik får man en bestämning av faktorladdningar, som är relativt invariant med hänsyn till olika person- och attributstickprov. Horst (1965, s 469) menar att ett sådant synsätt ur filosofiska synpunkter kan vara befogat. Men sett ur ett rent formellt perspektiv, där man enbart tar hänsyn till de matematiska operationerna i faktoranalysen, finns det inga skäl varför man skulle kunna vara mera intresserad av faktorladdningsmatrisen än av faktorvärdematrisen.

Vissa forskare har hävdat, att faktorvärdematrisen ej ger någon information utöver de mätvärden på vilka matrisen baseras. Och därför, menar man, är faktorvärdematriser enbart av teoretiskt intresse. Men tyvärr, menar Horst (1965, s 469), har dessa forskare ställt frågan fel. Om man ställer frågan på detta sätt, finner man, att datamatrisen kan ge upplysning, inte enbart med hänsyn till relevant eller systematisk information, utan också om slumpmässig och ej reliabel information.

På så sätt kan man nyttja den lägre rangapproximeringen av datamatrisen, dvs faktorvärdematrisen, som en metod för att eliminera slumpmässig eller irrelevant varians från datamatrisen. Den största fördelen med principalaxelmetoden är därvid att den ger samtidigt (1) den minsta kvadrat-, (2) en rangreducerings- och (3) en ortogonal-lösning för faktorvärdematrisen (Horst, 1965, s 476).

11. FAKTORANALYSENS TRE VANLIGASTE TILLÄMPNINGSFORMER

11.1 Att strukturera en datamängd

En faktor är ett sammanfattande påstående om en lineär relation mellan olika variabler i en variabeluppsättning. Denna relation kan matematiskt utnyttjas för att representera dessa variabler (Eysenck, 1967, s 290).

Denna definition är av rent formell natur och äger ej på något sätt ett psykologiskt innehåll eller en kausal implikation, ej heller hypoteser vare sig av prövade eller av uppslagsgivande natur. Det har ej funnits

många forskare på det psykologiska området, som på detta sätt ansett faktoranalysmetoder vara särskilt attraktiva. Exemplen på tillämpning av faktoranalytiska metoder enligt denna definition är fåtaliga. Påståendet att sådana faktorer är av ringa psykologiskt intresse anser Eysenck (1967, s 290) som en felaktig kritik av analyser, som på detta sätt utförts som rent deskriptiva. Att ha 5 faktorer i stället för t ex 35 variabler i en regressionsekvation rättfärdigar denna begränsade användningsform av faktoranalysmetoder.

11.2 Att formulera hypoteser

Hypoteser kan genereras på många olika sätt. Det finns ingen garanti för att en hypotes som bygger på faktorer är överlägsen sådana hypoteser som har formulerats mot bakgrund av t ex observationer, av en teoretisk analys eller av s k skrivbordsexperiment.

Ett exempel på en hypotes som bygger på en rent teoretisk analys är Sprangers (1930) människotyper: den teoretiska, ekonomiska, estetiska, sociala och religiösa personlighetstypen. En faktoranalytisk undersökning av Sprangers typer, utförd av Laurie (1951) resulterade i fyra faktorer: social, teoretisk, religiös och politisk (en kombination av ekonomiska och politiska intressen) samt en estetisk faktor. Undersökningen stödde alltså Sprangers hypotes. Faktoranalysen genomfördes för att testa denna hypotes, dvs analysen användes i hypotesprövande syfte.

Men denna typanalys kan också tjäna som exempel på formulering av hypoteser. Medan man alltså vanligen kan formulera hypoteser på många olika sätt, medför faktoranalystekniken en definitiv fördel. Den ger data, som underlättar formuleringen av en hypotes. Dessutom medför metoden förkastandet av ett stort antal hypoteser, som i annat fall skulle ha blivit föremål för resultatdiskussion. Som formell definition för faktoranalyser av denna typ skulle kunna accepteras: "En faktor är ett kondenserat påstående om lineära relationer bland variabler i en variabeluppsättning och ger uppslag till hittills ej upptäckta kausala relationer" (se Eysenck, 1967, s 291).

Denna användningsform är den vanligaste. Framför allt i explorativa studier kan faktoranalystekniken vara till mycket stor hjälp i syfte att snabbare komma fram till formuleringen av meningsfulla hypoteser och förkastande av sådana, som visar sig vara sämre alternativ.

11.3 Att pröva hypoteser

I viss mån har faktoranalysmetodens hypotesprövande aspekter redan berörts i samband med det som sagts om olika roteringstekniker. Varje roteringsteknik bygger nämligen på specifika antaganden, dvs hypoteser. Har man formulerat en hypotes (resp flera alternativa hypoteser för en och samma faktor), bör man med denna (resp dessa) som bas ställa upp deduktioner samt testa dessa mera specifikt i nya undersökningar.

Det är påtagligt att faktoranalysen ej kan användas som en formell del med hänsyn till hypotetiko-deduktiv-processer i relation till vilka hypotestyper som helst. Största delen av de hypoteser som ställs inom psykologin, kräver nämligen någon form av beroende analyser, som i sin tur medför att faktoranalysmetoden ej kan användas som analysinstrument.

Faktoranalysen lämpar sig däremot mycket bra för hypotestestning, när det gäller analysen av datamängdens struktur och organisation. Under denna kategori faller alla typ-.och egenskapshypoteser. Den formella definitionen är då:

En faktor är ett kondenserat påstående om lineära relationer bland variabler i en variabeluppsättning, som överensstämmer med prediktioner, baserade på en teoretisk analys (se Eysenck, 1967, s 292).

Den enklaste prediktionen är att vissa test eller variabler laddar antingen positivt eller negativt i en viss faktor, som man sedan prövar faktoranalytiskt.

Eysenck konstruerade t ex items för att mäta sociala attityder, som skulle kunna specificeras med en obetydlig felmarginal inom ramen av en definierad två-faktor-rymd.

Betraktar man en faktor eller faktorstruktur som en hypotetisk begreppsbildning med kausala implikationer och bestämning av relationen mellan variablerna i en variabeluppsättning, får man precis det som är av särskilt intresse i den psykologiska forskningen. Denna funktion medför samtidigt att faktoranalysens hypotesgenererande och hypotesprövande funktioner kommer mera i blickpunkten framför den rent deskriptiva funktionen. I en och samma undersökning förekommer det ofta att det finns faktorer som ger stöd för en viss hypotes och faktorer, som ger upphov till nya hypoteser.

Ett ytterligare användningssätt kan vara att undersöka om en ny variabeluppsättning överensstämmer med redan säkrade faktorer genom att man i sin analys tar med ett antal markeringsvariabler.

För att utesluta att dessa variabler eventuellt leder till en uppspaltning i nya gemensamma faktorer, utför man faktoranalysen enbart med hänsyn till de redan kända markeringsvariablerna.

Det är också möjligt att testa om det finns strukturlikheter i olika material. Man kan t ex faktoranalysera ett faktoriellt experiment i sin helhet och sedan undersöka om det finns strukturlikheter genom att jämföra helhetsanalysen med olika delanalyser avseende olika nivåkombinationer.

Strukturlikheter, när det gäller upprepade mätningar på samma individpopulation och med samma variabler (test) kan delvis undersökas med kanonisk korrelationsanalys. Rörande likhet i strukturen mellan olika oberoende individpopulationer (och samma variabler) finns inga bra test. Det minsta krav som man borde ställa på likhet är likhet i egenvärdesfördelningar.

Av de analysmetoder som sammanfattas under begreppet "latent strukturanalys" är faktoranalysen ett specialfall. Flera statistiker anser att såväl komponentanalysen som faktoranalysen för närvarande är metoder med tvivelaktiga värden (se Dempster, 1969, ss 139-140).

Det kan därför vara av vikt att än en gång poängtera: Resultatet av en faktoranalytisk undersökning beror i stor utsträckning på experimentatorns skarpsinnighet och förmåga att avgränsa och formulera hypoteser och att konstruera test samt på användingen av faktoranalystekniken som "hypotetiko-deduktivt" verktyg i stället för som en deskriptiv metod.

Men även metodvalet bestämmer delvis undersökningens resultat. När det gäller en rent deskriptiv användning, finns ej någon anledning till tvekan i det faktoranalytiska metodvalet. Metoder som t ex Hotellings, Tryons, Thurstones eller Kelleys är lika bra.

Först när man kommer in på valet av roteringsteknik blir metodvalet ett problem. Man väntar t ex ej, att en faktor är direkt relaterad till ett hypotetiskt orsakssamband, så länge denna faktor inte visar sig vara unik och invariant i förhållande till olika undersökningar och analysmetoder. Oroterade faktorer av vilken form som helst är vanligen varken unika eller invarianta. Vill vi alltså få faktorer, som har egenskapen att vara "allmänna" och invarianta, krävs det någon form av rotering.

Om vi kan behandla våra variabeluppsättningar som om kommunaliteten vore bestämd av ett litet antal isolerbara orsaker, skulle den bästa vägen vara att isolera och mäta dessa orsaker genom test- eller variabel-urvalet, vars varians är beroende av ett fåtal av dessa orsaker. Ett sådant förfaringssätt bör leda till klara och entydiga avgränsade faktorer. Ett sådant urval bör samtidigt kunna vara predicerbart.

12. TERMINOLOGI

Blindrotation	Primärlösningen av faktormatrisen (den oroterade) roteras med hjälp av ett internt kriterium som t ex varimaxkriteriet.
Datamatris	En systematiskt ordnad mängd av kvantifierade observationer, dvs värden för ett antal individer i ett antal variabler.
Dimension	Dimensioner är sammanfattande variabler, som bygger på en klassificering av de observerade variablerna med antagandet att de mäter samma sak. Sammanvägning eller direkt summering av variabelvärden sker till en ny uppsättning variabler som kallas dimensioner. Dimensioner är definierade genom den statistiska sammanvägningen.
Domän	Begreppet används ofta synonymt med population men endast avseende variabler.
Egenvärde	En faktors egenvärde är synonymt med dess varians. Summan av faktorernas egenvärdenär den varians som faktorerna tillsammans utvisar. Denna varians är lika med summan av variablernas kommunaliteter. Den del som en faktor utvisar av den totala variansen kan anges som ett proportionstal, dvs ett uttryck för andelen varians som en viss faktor utvisar.
Enkelstrukturkriterium	Enkel struktur (s 18) är en idealmodell, som man strävar att uppnå. Detta kan ske genom (1) grafisk transformation eller rotation eller (2) genom bestämning av statistiska kriterier.

Faktorladdning	Faktorladdningar är regressionsvikter för regressionen av de observerade variablerna på faktorerna. Faktorladdningarna är element i den roterade matrisen (faktorladdningsmatris). En faktorladdning kan ses som produkt-moment-korrelation mellan observerade variabler och skattade faktorer.
Faktorvärde (-poäng)	Individens värde på faktorn, regressionsskattning från de manifesta variablerna med hjälp av faktorladdningarna.
Felvarians	En variabels felvarians är restvariansen minus den specifika variansdelen. Felvariansen betraktas som slumpmässig, osystematisk och okorrelerad inom sig själv och i förhållande till övriga delar.
Gemensam varians	Begreppet är en teknisk term för den del i en variabel som förklaras av kovarians med en eller flera andra variabler, dvs den del av en variabels varians som är predicerbar från andra variabler.
Inferens	Vid definition av faktoranalys kan inferens antas ske till antingen individ- eller variabelpopulationer eller till båda.
Iteration	Ett matematiskt förfaringssätt för att stegvis beräkna närmevärden till en ekvationslösning.
Kommunalitet	Detta begrepp avser den del av en variabels varians som förklaras av faktorerna.
Kovarians	Den del av variabelvariansen som två variabler har gemensamt. Den kallas också samvarians.
Latenta variabler	Med latent avses ej direkt observerbar. I princip kan begreppet latent även innefatta ej direkt observerbara individpopulationer, men detta är en mindre vanlig användning.

Manifesta variabler	Med manifest avses direkt observerbara eller experimentella variabler.
Målrotering	Primärlösningen av faktormatrisen (den oroterade) roteras med hjälp av ett yttre kriterium som specificerar en hypotes om faktorstrukturen.
Population	Begreppet är av statistisk natur och avser den grupp av individer eller variabler man vill generalisera till.
Rang	Rang är en beteckning för matrisens maximala antal lineärt oberoende rader eller kolumner i matrisen. Denna kan vara rektangulär eller kvadratisk.
Referenssystem	Ur matematisk synpunkt är ett referenssystem enbart ett hjälpmedel, dvs ett matematiskt axelsystem för att t ex kunna göra tvådimensionella framställningar. Inom faktoranalysen har däremot referenssystem ett speciellt innehåll. Vid transformation av en faktormatris söker man ett nytt referenssystem som kan tolkas som en uppsättning dimensioner av meningsfullt innehåll.
Restvarians eller unik varians	En variabels restvarians eller unika varians är variabelns totala varians minus den gemensamma variansen. Denna variansdel är variabelspecifik. Den är inte predicerbar från andra variabler i datamatrisen. Variansdelen kan ytterligare spaltas upp i specifik varians och felvarians.
Specifik varians	Den del av restvariansen som inte är slumpvarians. Variabelns specifika varians kan inte predriceras.

Strukturanalys	Vid strukturproblem söker man en viss gruppering av de manifesta variablerna i en eller flera populationer. Även här ställer man frågan vad variablerna visar och förutsätter en reduktion av antalet variabler.
Systematisk varians	Den del av den totala variansen i en variabel som inte är slumpmässig. Variabelns gemensamma och specifika variansdel adderade ger variabelns systematiska varians. Proportionen systematisk varians i en variabel är enligt definitionen variabelns reliabilitet.
Varimax-metod	Denna metod är en blindroteringsmetod och utvecklades av Kaiser. Den innebär i princip ett försök att transformera de skattade faktorladdningarna till enkel struktur. Den har i olika sammanhang visat sig ge goda, tolkbara resultat.

13. REFERENSER

Burt, C. Experimental tests of general intelligence. Brit. J. Psychol., 1909, 3, 94-177.

Cattell, R. B. The description and measurement of personality. Yonkers, N. Y.: World Book Comp., 1946.

Cattell, R. B. Personality and motivation structure and measurement. Yonkers, N. Y.: World Book Comp., 1957.

Cattell, R. B. Theory of fluid and crystallized intelligence: A critical experiment. J. educ. Psychol., 1963, 54, 1-22.

Cattell, R. B. The three basic factor analytic research designs. - Their interrelations and derivatives. I: Jackson, D. N. & Messick, S. (Eds.) Problems in human assessment. New York, N. Y.: McGraw Hill, 1967, Ss. 300-304.

Child, D. The essentials of factor analysis. New York, Holt, Rinehart & Winston, 1970.

Coan, R. Basic form of covariation and concomitance designs. Psychol. Bull., 1961, 58, 317-324.

Cooley, W. W. & Lohnes, P. R. Multivariate procedures for the behavioral sciences. New York, N. Y.: Wiley & Sons, 1962.

Dempster, A. P. Elements of continuous multivariate analysis. Reading, Mass.: Addison-Wesley, 1969.

Eysenck, H. J. Structure of human personality. (2:a uppl.) London, Methuen, 1960.

Eysenck, H. J. The logical basis of factor analysis. I: Jackson, D. N. & Messick, S. (Eds.) Problems in human assessment. New York, N. Y.: McGraw Hill, 1967. Ss. 288-299.

Fruchter, B. Introduction to factor analysis. Princeton, N. J.: Van Nostrand, 1954.

Guilford, J. P. When not to factor analyze. I: Jackson, D. N. & Messick, S. (Eds.) Problems in human assessment. New York, N. Y.: McGraw Hill, 1967. Ss. 309-317.

Guttman, L. Multiple group methods for common-factor analysis: their basis, computation and interpretation. Psychometrika, 1952, 17, 209-222.

Guttman, L. Some necessary conditions for common-factor analysis. Psychometrika, 1954, 19, 149-161.

Harman, H. H. The square root method and multiple group methods of factor analysis. Psychometrika, 1954, 19, 39-55.

Horst, P. Multivariate models for evaluating change. I: Chester, H. (Ed.) Problems in measuring change. Madison, Wisc.: 1963. Ss. 104-121.

Horst, P. Factor analysis of data matrices. New York, N. Y.: Holt, Rinehart & Winston, 1965.

Hotelling, H. Analysis of a complex of statistical variables into principal components. J. educ. Psychol., 1933, 24, 417-441; 498-520.

Hotelling, H. The most predictable criterion. J. educ. Psychol., 1935, 26, 139-142.

Hotelling, H. Simplified calculation of principal components. Psychometrika, 1936, 1, 27-35.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 1958, 23, 187-200.

Kaiser, H. F. A note on the Tyron-Kaiser solution for the communalitics. Psychometrika, 1959, 24, 269-271.

Kelley, T. L. Essential traits of mental life. Harvard Stud. in Educ., 1935, 26. Cambridge, Mass.: Harvard Univ. Press.

Lurie, W. A. A study of Spranger's value-types by the method of factor analysis. J. soc. Psychol., 1951, 8, 17-37.

Morrison, D. F. Multivariate statistical methods. New York, N. Y.: McGraw Hill, 1967.

Mårdberg, B. Dimensionsanalys för beteendevetare. Stockholm: Personaladministrativa Rådet, meddelande nr 60, 1969.

Pawlik, K. Dimensionen des Verhaltens. Eine Einführung in Methodik und Ergebnisse faktorenanalytischer psychologischer Forschung. Stuttgart: Huber, 1968.

Royce, J. R. Factors as theoretical constructs. I: Jackson, D. N. & Messick, S. (Eds.) Problems in human assessment. New York, N. Y.: McGraw Hill, 1967, Ss. 318-325.

Rozeboom, W. W. Foundations of the theory of prediction. Homewood, Ill.: Dorsey Press, 1966.

Sjöberg, L. Factor analysis and multidimensional scaling, del I och II. Göteborg Department of Psychology. University of Göteborg, Sweden, 1970.

Spranger, E. Lebensformen. Geisteswissenschaftliche Psychologie und Ethik der Persönlichkeit. (7. Aufl.) Halle (Saale): Max Niemeyer Verlag, 1930.

Thomsen, G. The factorial analysis of human ability. (5. edition). London: University Press, 1951.

Thurstone, L. L. Multiple factor analysis. Psychol. Rev., 1931, 38, 406-427.

Thurstone, L. L. Multiple factor analysis. Chicago: Univ. of Chicago Press, 1947.

Thurstone, L. L. The factor problem. I: Jackson, D. N. & Messick, S. (Eds.) Problems in human assessment. New York, N. Y.: McGraw Hill, 1967, Ss. 279-287.

Tryon, R. C. So-called group factors as determiners of ability. Psychol. Rev., 1932, 39, 403-439.

Tucker, L. R. Implications of factor analysis of three way matrices for measurement of change. I: Chester, H. (Ed.) Problems in measuring change. Madison, Wisc.: 1963, Ss. 122-137.

Tucker, L. R. The objective definition of simple structure in linear factor analysis. I: Jackson, D. N. & Messick, S. (Eds.) Problems in human assessment. New York, N. Y.: McGraw Hill, 1967. Ss. 305-308.